

Anderson D, Cordell HJ, Fakiola M, Francis RW, Syn G, Scaman ESH, Davis E, Miles SJ, McLeay T, Jamieson SE, Blackwell JM.

[First Genome-Wide Association Study in an Australian Aboriginal Population Provides Insights into Genetic Risk Factors for Body Mass Index and Type 2 Diabetes.](#)

PLoS One 2015, 10(3), e0119333

Copyright:

© 2015 Anderson et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

DOI link to article:

<http://dx.doi.org/10.1371/journal.pone.0119333>

Date deposited:

01/05/2015



This work is licensed under a [Creative Commons Attribution 4.0 International License](#)

RESEARCH ARTICLE

First Genome-Wide Association Study in an Australian Aboriginal Population Provides Insights into Genetic Risk Factors for Body Mass Index and Type 2 Diabetes

Denise Anderson¹, Heather J. Cordell², Michaela Fakiola^{1,3}, Richard W. Francis¹, Genevieve Syn¹, Elizabeth S. H. Scaman¹, Elizabeth Davis^{1,4}, Simon J. Miles⁵, Toby McLeay⁵, Sarra E. Jamieson¹, Jenefer M. Blackwell^{1,3*}

1 Telethon Kids Institute, The University of Western Australia, Subiaco, Western Australia, 6008, Australia, **2** Institute of Genetic Medicine, Newcastle University, Newcastle upon Tyne, NE1 3BZ, United Kingdom, **3** Cambridge Institute for Medical Research, Department of Medicine, and Department of Pathology, University of Cambridge, Cambridge, United Kingdom, **4** Department of Endocrinology and Diabetes, Princess Margaret Hospital for Children, Subiaco, Western Australia, 6008, Australia, **5** Ngangganawili Aboriginal Health Service, Wiluna, Western Australia, 6646, Australia

* These authors contributed equally to this work.

* jenefer.blackwell@telethonkids.org.au



OPEN ACCESS

Citation: Anderson D, Cordell HJ, Fakiola M, Francis RW, Syn G, Scaman ESH, et al. (2015) First Genome-Wide Association Study in an Australian Aboriginal Population Provides Insights into Genetic Risk Factors for Body Mass Index and Type 2 Diabetes. PLoS ONE 10(3): e0119333. doi:10.1371/journal.pone.0119333

Academic Editor: Shiro Maeda, Graduate School of Medicine, University of the Ryukyus, JAPAN

Received: August 2, 2014

Accepted: January 28, 2015

Published: March 11, 2015

Copyright: © 2015 Anderson et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: De-identified genotype and basic demographic data (broad geographical location, age, sex and phenotype information) are available through the European Genome-phenome Archive (accession number EGAS00001001004) and the associated study-specific Data Access Committee.

Funding: Supported by the Australian National Health and Medical Research Council (APP634301). HJC is supported by a Senior Research Fellowship in Basic Biomedical Science from the Wellcome Trust

Abstract

A body mass index (BMI) $>22\text{kg/m}^2$ is a risk factor for type 2 diabetes (T2D) in Aboriginal Australians. To identify loci associated with BMI and T2D we undertook a genome-wide association study using 1,075,436 quality-controlled single nucleotide polymorphisms (SNPs) genotyped (Illumina 2.5M Duo Beadchip) in 402 individuals in extended pedigrees from a Western Australian Aboriginal community. Imputation using the thousand genomes (1000G) reference panel extended the analysis to 6,724,284 post quality-control autosomal SNPs. No associations achieved genome-wide significance, commonly accepted as $P < 5 \times 10^{-8}$. Nevertheless, genes/pathways in common with other ethnicities were identified despite the arrival of Aboriginal people in Australia $>45,000$ years ago. The top hit (rs10868204 $P_{\text{genotyped}} = 1.50 \times 10^{-6}$; rs11140653 $P_{\text{imputed}_1000\text{G}} = 2.90 \times 10^{-7}$) for BMI lies 5' of *NTRK2*, the type 2 neurotrophic tyrosine kinase receptor for brain-derived neurotrophic factor (BDNF) that regulates energy balance downstream of melanocortin-4 receptor (MC4R). PIK3C2G (rs12816270 $P_{\text{genotyped}} = 8.06 \times 10^{-6}$; rs10841048 $P_{\text{imputed}_1000\text{G}} = 6.28 \times 10^{-7}$) was associated with BMI, but not with T2D as reported elsewhere. BMI also associated with *CNTNAP2* (rs6960319 $P_{\text{genotyped}} = 4.65 \times 10^{-5}$; rs13225016 $P_{\text{imputed}_1000\text{G}} = 6.57 \times 10^{-5}$), previously identified as the strongest gene-by-environment interaction for BMI in African-Americans. The top hit (rs11240074 $P_{\text{genotyped}} = 5.59 \times 10^{-6}$, $P_{\text{imputed}_1000\text{G}} = 5.73 \times 10^{-6}$) for T2D lies 5' of *BCL9* that, along with *TCF7L2*, promotes beta-catenin's transcriptional activity in the WNT signaling pathway. Additional hits occurred in genes affecting pancreatic (*KCNJ6*, *KCNA1*) and/or GABA (*GABRR1*, *KCNA1*) functions. Notable associations observed for genes previously identified at genome-wide significance in other populations included *MC4R* ($P_{\text{genotyped}} = 4.49 \times 10^{-4}$) for BMI and *IGF2BP2* ($P_{\text{imputed}_1000\text{G}} =$

(Grant references 087436/Z/08/Z and 102858/Z/13/Z). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

2.55×10^{-6}) for T2D. Our results may provide novel functional leads in understanding disease pathogenesis in this Australian Aboriginal population.

Introduction

Genome-wide association studies (GWAS) have been used with great success to identify genes associated with complex diseases [1], including obesity and type 2 diabetes (T2D) [2–6]. However, there are no published data on the use of this approach to study complex diseases in Aboriginal Australians. This is partly a reflection of the controversy surrounding genetic research in indigenous communities [7, 8], which has raised ethical concerns including a lack of benefit to community and diversion of attention and resources from non-genetic causes of health disparities and racism in health care [9–12]. Controversy relating to the Human Genetic Diversity Project has acted as a particular barrier to conducting genetic research in Australian Aboriginal communities [13], but there is now a strong lobby to bring the benefits of health-based genomic research to Australian Aboriginal populations [14] as there is for all ethnicities globally [8, 15]. T2D and associated pathologies are a major health problem in Indigenous Australians, and a body mass index (BMI) $>22\text{kg/m}^2$ is a significant risk factor for T2D in Australian Aboriginal populations [16]. Here we present the first GWAS of BMI and T2D in a Western Australian Aboriginal population.

Materials and Methods

Study population

Subjects for the discovery GWAS were recruited from an Australian Aboriginal community of Martu ancestry [17, 18] at the edge of the Western Desert in Western Australia. A family-based study was proposed to take account of recent population history and shared ancestry. A memorandum of understanding (MoU) was established with the community, which included permission for access to clinical records held in a Communicare database at the local Aboriginal Health Service. An individual was classified as having T2D if the subject was: (1) diagnosed with T2D by a qualified physician; (2) on a prescribed drug treatment regimen for T2D; and (3) returned biochemical test results of a fasting plasma glucose level of at least 7 mmol/l in SI units based on criteria laid down by the World Health Organization (WHO) consultation group report [19]. Multiple height and weight measures per individual were recorded in Communicare through time and converted within the database to BMI according to the standard formula $\text{BMI} = \text{weight (kg)}/\text{height (m)}^2$. DNA was prepared from saliva samples collected into Oragene tubes (DNA Genotek, Ontario, Canada) from 405 consenting family members who were available at the time of visits by the study team during the two-year collection period of the study.

Ethical approvals

Ethical approval for the study was obtained from the Western Australian Aboriginal Health Ethics Committee (WAAHEC; Reference 227 12/12), who reviewed and approved forms for informed consent. Each individual (or the parent or guardian of individuals less than 18 years of age) signed separate informed consent forms to participate in the study and to provide a DNA sample. Following feedback of results to the community, permission to publish was provided by the Board of the local Aboriginal Health Service, which comprised elders representing the extended families residing in the area. Permission to lodge de-identified genotype and

basic demographic data (broad geographical location, age, sex and phenotype information) in the European Genome-phenome Archive (accession number EGAS00001001004) was also obtained from the Board of the local Aboriginal Health Service.

Genotyping, quality control, and analysis of population structure

DNA from the 405 consenting individuals were genotyped on the Illumina Omni2.5 BeadChip (outsourced to the Centre for Applied Genomics, Toronto, Ontario, Canada). Quality control (QC) data from the service provider indicated a call rate of (mean [SD]) 99.680% [0.005%]. Further in-house QC procedures at the level of individuals resulted in one individual being dropped due to a missing data rate >5%, 2 exclusions due to unintentional duplication, and no exclusions due to outlying heterozygosity (using 3 standard deviation limits), discordant sex, or divergent ancestry (since mixed models were employed to take account of ancestry, cf. below). This provided a post-QC dataset of 402 family members, including 361 with repeated body mass index measurements and 391 (89 cases, 302 family members unaffected at the time of collection) with information on doctor diagnosed T2D as per criteria outlined above. SNPs with minor allele frequency (MAF) <0.05, or with more than 5% missing data, were removed prior to association analysis (cf. below). As family data were employed in our study, complicated by a degree of over-relatedness in the pedigrees, a check of Hardy Weinberg Equilibrium (HWE) was not used as a SNP QC check since it was not clear that we should expect all SNPs to be in HWE. A subset of 70,420 genotyped SNPs with pairwise linkage disequilibrium (LD; r^2) ≤ 0.3 and MAF >0.01 was used in principal component analysis (PCA; SMARTPCA within EIGENSOFT [20, 21]) to look at population substructure across the 402 family members.

Imputation procedures

Pre-phasing of genotyped data was performed using SHAPEIT [22], and imputation conducted using IMPUTE2 [23] with 1000 Genomes (1000G) haplotypes [Phase I integrated variant set release (v3)]. The reference panel includes haplotypes of 1,092 individuals from Africa, Asia, Europe and the Americas. For association analysis, probabilistic imputed SNP genotypes were converted to hard genotype calls provided the posterior probability of the most likely genotype was >0.9, and imputed SNPs were retained only if they had 'info' score >0.5, MAF>0.01 and <10% missing data. To assess imputation accuracy, IMPUTE2 provides measures of concordance and squared Pearson correlation (r^2) for each genotyped SNP. Directly genotyped SNPs were masked then imputed and concordance between the true genotype and the most likely imputed genotype [where probability prediction threshold > 0.9] for each SNP was calculated [24–26], as well as r^2 between the true (discrete) allele dosage (coded as 0, 1 and 2 corresponding to the number of minor alleles) and the imputed (continuous) allelic dosages. For low frequency SNPs, r^2 is the preferred accuracy measure because, unlike concordance, allele frequency is taken into account. We also compared these imputation accuracies for the 195 individuals determined by PCA to represent pure Martu ancestry against the 207 deemed to be of mixed ethnicity by PCA, using the data generated from imputation of the full dataset for the 402 individuals.

Heritability and association analyses

Heritability for mean BMI was determined from self-reported pedigree structures using the QTDT software package [27], and separately on the basis of kinship and IBD sharing from SNP-chip data calculated using the R package GenABEL v1.7–6 [28] and using the genome-wide complex trait association (GCTA) tool [29]. To provide a visual display of relatedness across the 402 genotyped individuals, the genomic kinship matrix was first calculated in

GenABEL v1.7–6 [28], and then converted to a distance matrix followed by hierarchical cluster analysis using single linkage on the dissimilarities. The ape package [30] was used to produce a radial tree plot for the 402 genotyped individuals. For association analyses, the FAmily-based Score Test for Association (FASTA) [31] in GenABELv1.7–6 [28] was employed that uses a linear mixed model (LMM) approximation to model the trait outcome, with whole genome data used to estimate kinship (in order to account for relatedness) and to take account of population substructure. Analyses were carried out using quantitative trait data (BMI), and separately to compare all T2D cases with all unaffected individuals. Association results for genotyped data obtained using GenABEL were compared with results using FaST-LMM [32]. Both GenABEL (FASTA) and FaST-LMM model disease status (control/case for T2D, coded 0/1) as if it were a normally distributed quantitative variable, which has been shown [33] to produce a valid test with respect to testing the null hypothesis of no association. Power calculations using QUANTO [34] show that 361 individuals with BMI measures provide a maximum of 80% power to detect associations at genome wide significance levels ($P < 5 \times 10^{-8}$), or a maximum of 97% power at suggestive significance levels ($P < 1 \times 10^{-5}$), for SNPs with MAF > 0.3 conferring allelic effects (betas) of magnitude half a unit of standard deviation. For effects (betas) of magnitude 0.25 units of standard deviation, the maximum powers achievable are lowered to 1% and 9% respectively. The 89 T2D cases and 302 unaffected family members provide a maximum of 32% power to detect associations at genome wide significance levels ($P < 5 \times 10^{-8}$), or 71% power at suggestive significance levels ($P < 1 \times 10^{-5}$), for SNPs with MAF > 0.3 conferring allelic odds ratios > 2.5. For odds ratios of 1.5, the powers are lowered to 1.4% and 0.06% respectively.

Manhattan plots were generated using the mhtplot() function of 'gap', a genetic analysis package for use in R (see URLs). Quantile-quantile (Q-Q) plots were generated, and inflation factors (denoted λ) calculated in R version 2.15.0 by dividing the median of the observed chi-squared statistics by the median of the theoretical chi-squared distribution. Regional plots of association were created using LocusZoom [35] in which $-\log_{10} P$ -values were graphed against their chromosomal location. Pairwise LD patterns between all regional SNPs and the top SNP were calculated specifically for this Australian Aboriginal study data using 146 unrelated individuals from the total sample of 402 genotyped individuals. The 146 unrelated individuals were selected by iterative removal of the individuals with the greatest number of estimated relationships with IBD > 0.1875.

Bioinformatic analysis

Global alignment of genomic sequence for the region *SLC28A3* to *NTRK2* from human, cow and mouse was undertaken to locate evolutionarily conserved non-coding sequences (CNS) that might contain regulatory elements and transcription factor binding sites (TFBS). Genomic sequences for the three organisms were exported in FASTA format from ENSEMBL (Genome Reference Consortium Release 37, Ensembl Release 67) and associated annotation exported in the form of a General Feature File (GFF) file. The global alignment tool Multi-LAGAN [36, 37] was used to align genomic sequences using the guide tree ((human) cow) mouse). SYNPLLOT [38] was used to visualize the annotated alignment. CNS, defined here as regions with a nucleotide sequence conservation level of ≥ 0.7 (i.e. \geq to the least conserved exon sequence in the genes flanking the region of interest), were analysed for promoter and enhancer elements using PROMO [39], AliBaba v2.1 [40], and MatInspector v8.0.5 [41] with a matrix similarity parameter > 0.75. We also used the UCSC genome-browser (see URLs) with custom tracks to assess where selected top-hit SNPs are located in relation to elements such as CpG islands or repeat elements (LINES; SINES). To look *de novo* for CpG island-like elements, we firstly masked repeat regions using RepeatMasker (see URLs) following which putative CpG islands were searched

for using CpGIsland Searcher [42] with parameters set at 50%GC, 0.60 observed CpG/expected CpG ratio, and length 200bp (rather than the default settings of 55%, 0.65, and 500bp).

Results

Characteristics of the study population

The 402 post-QC genotyped individuals used in the GWAS belonged to a small number of inter-related extended pedigrees, as depicted in the radial plot which shows hierarchical clustering of estimated pairwise identity-by-descent allele-sharing (S1 Fig.). Principal component analysis (S2 Fig.) demonstrated a degree of introgression of predominantly Caucasian origin, with a tight cluster of 195 individuals of Martu Aboriginal ancestry across all age groups. Linear mixed models were used in the genetic analysis to take account of both family relationships and this genetic substructure. S1 Table provides basic demographic data (age, sex; at the time of diagnosis and/or collection) for the 391 individuals used in the genetic analyses. De-identified BMI data was also available for 1020 self-reporting Aboriginal individuals in the population-specific Communicare database.

BMI in the study population

A BMI $>22\text{kg/m}^2$ is a significant risk factor for T2D in Australian Aboriginal populations [16]. Fig. 1 part A provides a plot of all BMI measurements for self-reported Aboriginal individuals in the study population ($N = 1020$). BMI by age was plotted using the R package SITAR [43] for all records in Communicare. Separate lines trace multiple measurements over time per individual; females (pink lines) and males (blue lines). The heavy lines show the polynomial quintic (power of 5) curves for females (pink heavy line) and males (blue heavy line) that best fit the data. Fitting separate (by gender) curves to the data (Fig. 1 part A) provided a significantly better fit ($P = 10^{-9}$) than fitting a single curve with a sex-specific displacement. We constructed standardized residuals from these curves by subtracting the observed BMI value from the fitted curve and dividing by the estimated standard deviation (calculated within 1-year age brackets). All GWAS analyses of BMI (cf. below) were carried out using these standardized residuals. The extreme female outlier was not used in fitting the female polynomial curve or in any of the GWAS analyses. This analysis of BMI shows (a) that the majority of the adult population have BMI >22 and are therefore at increased risk of diabetes; (b) a proportion of 15–20 year olds also fall into this category; and (c) there are many individuals in the community who can be classified as overweight (BMI 25.00–29.99) or obese (including class I obesity BMI 30.00–34.99; severe or class II obesity BMI 35–39.99; and morbid or class III obesity BMI ≥ 40) according to Center for Disease Control (CDC) [44] and World Health Organization (WHO) [45] international criteria. By estimating familial correlations based on self-reported pedigree structures using QTDI [27], we estimated heritability for mean BMI in this population to be 55%, broadly in line with other populations internationally [46–48]. Estimation of heritability on the basis of kinship and IBD sharing from the SNP-chip data using the R package GenABEL [28] gave a lower estimate of 38%, consistent with the estimate of 39% (95% confidence interval [17%–61%]) provided by GCTA [29].

T2D in the study population

Our dataset for the study population (Fig. 1 part B) showed that $>75\%$ of adults (≥ 20 years of age) fall above the BMI cut-off of >22 for increased risk of T2D in Aboriginal Australians [16]. Fig. 1 part B also highlights BMI curves for individuals in the study sample with T2D (heavy pink lines for females; heavy blue lines for males). This demonstrates that our study population

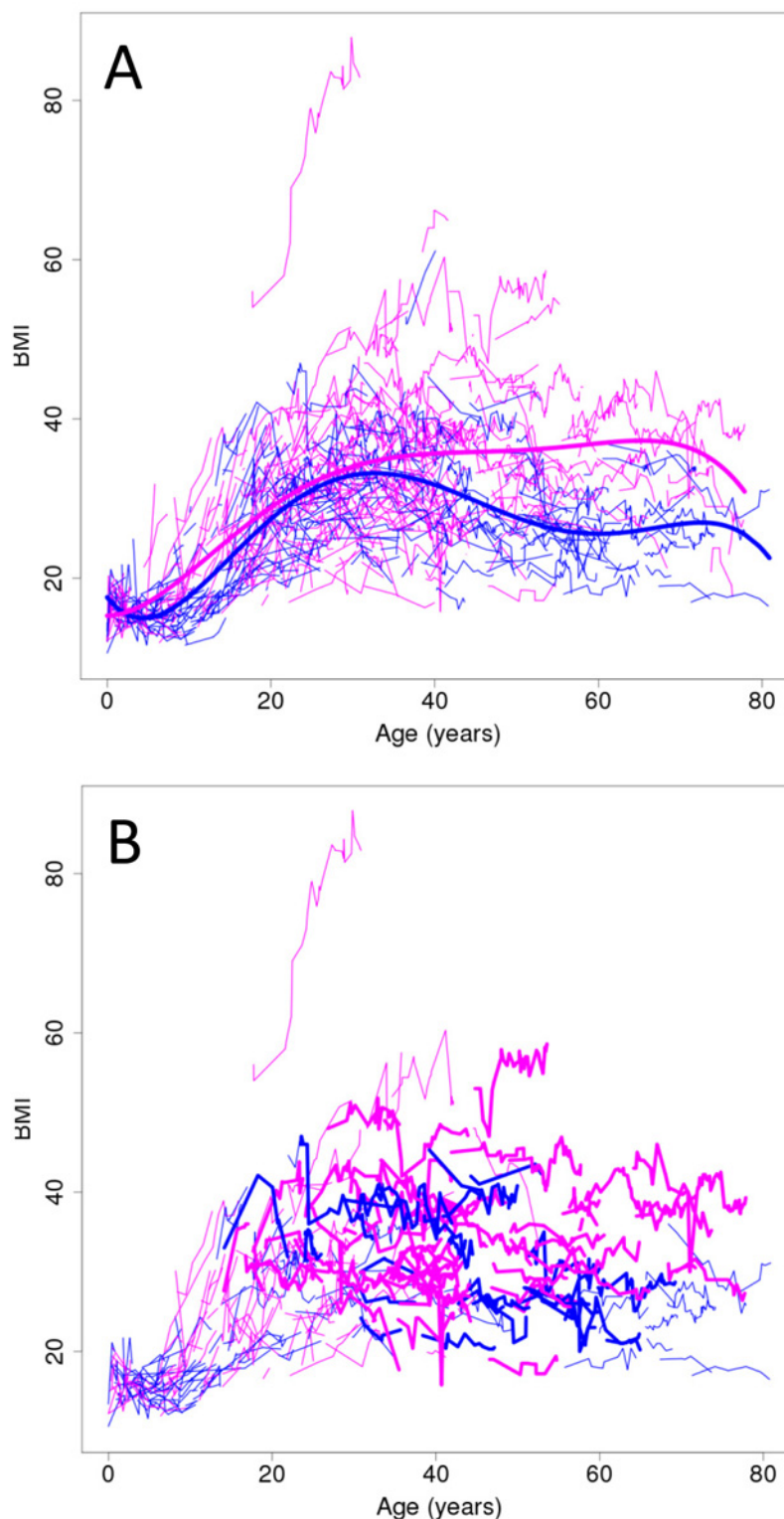


Fig 1. BMI by age plotted using the R package SITAR. (A) plot of all records for self-reporting Aboriginals ($N = 1020$) in the Aboriginal Health Service's Communicare database; and (B) plot of all records for the 391 genotyped individuals contributing to association analyses for BMI and T2D. Separate lines trace multiple measurements over time per individual; females (pink) and males (blue). The heavy lines in (A) show the polynomial quintic (power of 5) curves for females (pink heavy line) and males (blue heavy line) that best fit

the data. Fitting separate (by gender) curves to the data provided a significantly better fit ($P = 10^{-9}$) than fitting a single curve. The extreme outlier was not used in fitting the female polynomial curve. In (B), individuals with T2D are shown in heavy lines.

doi:10.1371/journal.pone.0119333.g001

is characterized by T2D predominantly associated with high risk BMI measurements, including a number of cases <20 years of age. [S2 Fig.](#) part D highlights on a PCA plot the individuals with T2D used in the GWAS. Of the 391 individuals used in the T2D GWAS (cf. below), there were 65 with T2D from 191 individuals (34%) that fell within the pure Martu ancestry cluster ([S2 Fig.](#) parts E and G), and 24 with T2D from 200 individuals (12%) of mixed ethnicity. There was insufficient power to analyse GWAS data separately for T2D associated with low to normal (<25; $N = 10$) BMI or with low (<20 years of age; $N = 5$) age at onset. There will be some loss of power in the analysis with the inclusion of unaffected individuals less than 20 years of age (since ~10–20% of them may go on to get T2D [[16](#)]). This had to be balanced against the gain in power with the increased sample size and greater contribution of individuals in controlling for population substructure and relatedness in the analysis. [S1 Table](#) shows the age range and BMI for unaffected individuals <20 and >20 years of age separately. Importantly, there will be no false positive associations generated by including all unaffected individuals in the GWAS analysis for T2D (cf. below).

GWAS analyses for BMI based on genotyped SNPs

Longitudinal BMI data available for the 361 post-QC individuals with both BMI measurements and genotyped SNP-chip data were representative of the full range of BMI values in the population ([Fig. 1](#) part B). To identify loci associated with BMI we undertook a GWAS using 1,075,436 post quality control genotyped SNPs in these 361 individuals. For longitudinal BMI data (designated BMI-longitudinal), the analyses in GenABEL and FaST-LMM used each individual BMI reading as a separate observation, using the standardized residual from the fitted curve as the trait of interest and modelling the correlation between readings via the estimated kinship. This does not fully account for the correlation between readings as readings from the same person are likely to be more correlated, especially since they are often measured at frequent (close together) time points. The Genomic Control deflation factor [[49](#)] was therefore employed in GenABEL to avoid inflation of the overall distribution of test statistics; this correction was not found to be necessary in FaST-LMM, as previously noted [[50](#)]. Association analyses were also undertaken in GenABEL and FaST-LMM using the mean of all BMI readings for an individual (designated BMI-mean) as the trait. Although this raises issues of differential variation in BMI-mean due to uneven numbers of readings between individuals, results from these two different analyses were well correlated. The plots presented at [S3 Fig.](#) demonstrate that association results for BMI-mean and BMI-longitudinal obtained using FASTA in GenABEL ([S3 Fig.](#) part A) or using FaST-LMM ([S3 Fig.](#) part B) were highly correlated. Therefore detailed results are given only for BMI-longitudinal (hereinafter referred to as BMI). Similarly, results obtained for each phenotype from GenABEL and FaST-LMM were strongly concordant ([S4 Fig.](#)), consistent with our recent evaluation of a range of software implementations for linear mixed models [[50](#)]. Q-Q plots ([S5 Fig.](#)) for GenABEL (λ BMI-longitudinal = 1.0; λ BMI-mean = 1.02) and FaST-LMM (λ BMI-longitudinal = 1.00; λ BMI-mean = 1.00) analyses also showed equivalent inflation factors. Therefore detailed results presented hereafter are given only for analyses undertaken using FASTA in GenABEL.

A Manhattan plot showing the genome-wide results for BMI based on the genotyped data is presented in [Fig. 2](#) part A. The top hits did not achieve genome-wide significance, commonly accepted as $P < 5 \times 10^{-8}$ [[51](#)] and concordant with the number of post-QC SNPs ($P = 0.05/$

1,075,436 or 4.65×10^{-8}) used for this analysis of genotyped data. No specific SNPs reported to achieve genome-wide significance for association with BMI in other populations achieved $P < 10^{-2}$ in our study, as determined by interrogation of the NIH NHGRI Catalogue of GWAS studies [52]. Nevertheless, since we cannot assume similar patterns of linkage disequilibrium or directionality of associations in our Australian Aboriginal population, we have provided information in S2 Table on SNPs with $P < 10^{-2}$ in our study population for the genes previously reported to achieve genome-wide significance (i.e. $P < 5 \times 10^{-8}$) for association with BMI in other populations. Notably (cf. below), *MC4R* (top SNP rs129959775; $P_{\text{genotyped}} = 4.49 \times 10^{-4}$) was present in this list. Extending the table to include genes previously reported in GWAS for BMI at $P < 10^{-5}$ in other populations provided evidence for association at *CNTNAP2* (top SNP rs6960319; $P_{\text{genotyped}} = 4.65 \times 10^{-5}$). SNPs at *CNTNAP2* were also among the top 50 SNPs from the BMI analysis in our study population (S3 Table). Data for *CNTNAP2* and other top SNPs in genes of potential functional relevance, i.e. in pathways previously identified to be associated with metabolic diseases, are shown in Table 1. Effect sizes (betas 0.35 to 0.72) for these associations are concordant with our power to detect allelic effects of magnitude half a unit of standard deviation of the trait. The top hit (rs10868204; $P_{\text{genotyped}} = 2.73 \times 10^{-6}$) for BMI in our study population lies in the intergenic region (Fig. 3; cf. below) 5' of *NTRK2* encoding the type 2 neurotrophic tyrosine kinase receptor for brain-derived neurotrophic factor (BDNF) that regulates energy balance downstream of melanocortin-4 receptor (*MC4R*). Both *BDNF* and *MC4R* have previously been shown to be associated with obesity in other populations (see S2 Table). Other top hits in our population that are of functional interest (Table 1) occurred in *RBM7* (rs6848632 $P_{\text{genotyped}} = 1.43 \times 10^{-5}$) which has been related to pancreatic function [53], and at *PIK3C2G* (rs12816270 $P_{\text{genotyped}} = 8.06 \times 10^{-6}$) which has previously been associated with T2D [54].

GWAS analyses for T2D based on genotyped SNPs

A Manhattan plot showing the GWAS results for T2D for genotyped SNPs is presented in S6 Fig. part A. Again, the top hits based on analysis of genotyped SNPs did not achieve genome-wide significance for T2D, which was less well-powered than the BMI analysis. In addition, no specific SNPs reported to achieve genome-wide significance for association with T2D in other populations achieved $P < 10^{-2}$ in our study, as determined by interrogation of the NIH NHGRI Catalogue of GWAS studies [52]. Nevertheless, as for BMI we cannot assume similar patterns of linkage disequilibrium or directionality of associations in our Australian Aboriginal population. Therefore, we have provided information in S4 Table for genotyped SNPs with $P < 10^{-2}$ in our study population for the genes previously reported to achieve genome-wide significance (i.e. $P < 5 \times 10^{-8}$) for association with T2D in other populations. Five genes (*ANK1*, *TSPAN8*, *PROX1*, *GLIS3* and *UBE2E2*) had genotyped SNPs with $P < 10^{-3}$. Of note, no SNPs at $P < 10^{-2}$ were observed for *TCF7L2* [55–59] or *KCNJ11* [60], genes that have previously been associated with T2D across multiple ethnicities. However, 4 genes (Table 2) represented in the top 50 SNPs (S5 Table) are supported by strong biological candidacy in related gene pathways. The top hit (rs11240074 $P_{\text{genotyped}} = 5.59 \times 10^{-6}$) for T2D in our study population lies 5' of *BCL9* encoding a protein that, along with *TCF7L2*, promotes beta-catenin's transcriptional activity in the WNT signaling pathway. Additional hits ($1.07 \times 10^{-4} \leq P_{\text{genotyped}} \leq 4.55 \times 10^{-5}$) of functional interest occurred in genes involved in pancreatic (*KCNJ6* [61], *KCNA1* [62]) and/or GABA (*GABRR1* [63], *KCNA1* [64]) functions.

GWAS analyses based on imputed SNP data

Recent reports have focused on improving tools for imputation of SNPs based on both the 1000 genomes (1000G) and HapMap project data [23, 65]. It was of particular interest to determine how well 1000G imputation would work for this Australian Aboriginal population, given

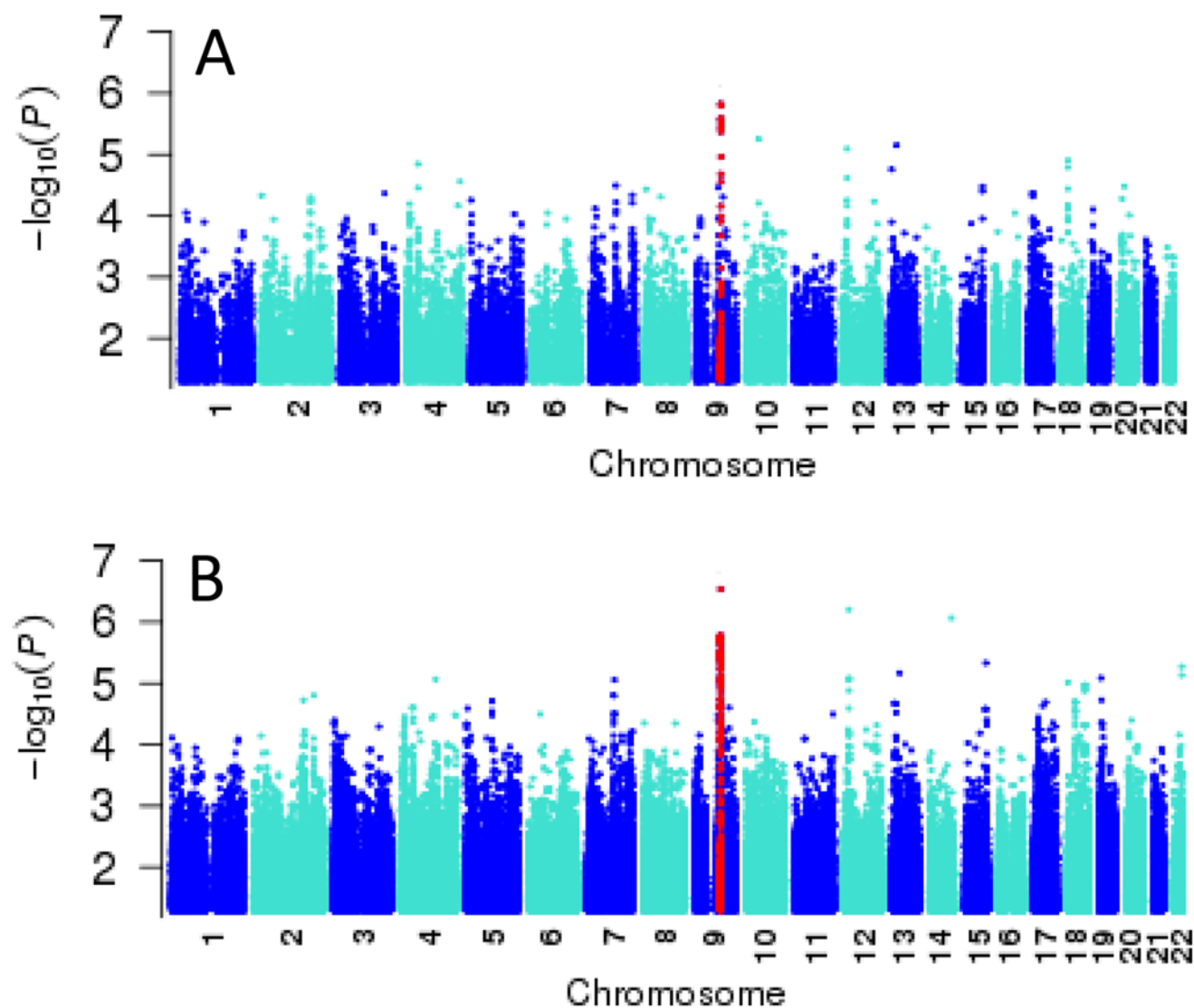


Fig 2. Manhattan plots of genome-wide association results for BMI undertaken using FASTA in GenABEL. (A) results for genotyped data; and (B) results for 1000G imputed data. SNPs in red show the region of the top association in this discovery GWAS.

doi:10.1371/journal.pone.0119333.g002

that most estimates suggest the arrival of Aboriginal people in Australia more than 45,000 years ago [66, 67]. We therefore examined the efficiency of imputing genotypes across the genomes of our study population using 1000G data. High average concordance (92.8%-97.9%) was observed across all chromosomes independently of MAF. Fig. 4 compares the efficiency of imputation across all chromosomes, as measured by imputation r^2 , for SNPs at different MAF. As expected, r^2 is lower for SNPs with MAF 0.01–0.03 (mean r^2 77.2–85.4) compared to SNPs with MAF 0.03–0.05 (mean r^2 81.5–88.5%) and MAF 0.05–0.5 (mean r^2 85.9–90.7%). When examined across individuals, mean imputation accuracy for the 195 individuals of pure Martu ancestry (concordance 94.8; r^2 88.8) were not different to measures for the 207 individuals of mixed ethnicity (concordance 95.6; r^2 90.0) or to measures across all 402 individuals (concordance 95.2; r^2 89.5).

Table 1. Top GWAS SNP hits in genes of functional relevance for BMI, organized by chromosome.

Chromosome	SNP	NCBI37	A1	A2	BMI Genotyped			BMI Imputed 1000G			SNP Location	HGNC*
					effB	se_effB	P _{deflated}	effB	se_effB	P _{deflated}		
4	rs6848632	40509521	A	C	0.52	0.08	1.43E-05				INTRONIC	RBM47
	rs6848632	40509521	A	C				0.51	0.08	2.50E-05	INTRONIC	RBM47
7	rs6960319	147258631	G	A	-0.36	0.06	4.65E-05				INTRONIC	CNTNAP2
	rs13225016	147256110	G	A				-0.35	0.06	6.57E-05	INTRONIC	CNTNAP2
9	rs1347857	87173097	A	C	0.44	0.06	1.50E-06				INTERGENIC	SLC28A3/NTRK2
	rs11140653	87152509	C	T				0.48	0.06	2.90E-07	INTERGENIC	SLC28A3/NTRK2
12	rs12816270	18770577	A	G	0.72	0.11	8.06E-06				INTRONIC	PIK3C2G
	rs10841048	18779259	T	C				0.71	0.10	6.28E-07	INTRONIC	PIK3C2G

Results are for GWAS analysis in GenABEL using each individual BMI reading as a separate observation, modelling the correlation between readings via the estimated kinship and using the Genomic Control deflation factor to avoid inflation of the overall distribution of test statistics. Results are for allele-wise tests under an additive model of inheritance for genotyped SNPs and for imputed data. Bold indicates the top hit for BMI based on both genotyped and 1000G imputed data. Full lists of the top 50 hits for genotyped SNPs, and the top 100 SNPs for imputed data, appear in [S3 Table](#) and [S6 Table](#), respectively.

* Genes separated by forward slash indicate nearest protein coding genes upstream/downstream of the SNP. NCBI37 = bp location on chromosome for NCBI Build 37. A1 = major allele; A2 = minor allele.

doi:10.1371/journal.pone.0119333.t001

In relation to the GWAS analyses for BMI ([Fig. 2](#) part B; [Fig. 3](#) part B; [Table 1](#); [S6 Table](#)), 1000G imputation improved significance for associations at *NTRK2* ([Fig. 3](#) part B; [Table 1](#): $P_{\text{genotyped}} = 1.50 \times 10^{-6}$; $P_{\text{imputed}_1000G} = 2.90 \times 10^{-7}$) and at *PIK3C2G* ([Table 1](#): $P_{\text{genotyped}} = 8.06 \times 10^{-6}$; $P_{\text{imputed}_1000G} = 6.28 \times 10^{-7}$) but not at other genes of functional interest ([Table 1](#)). Similarly, for the T2D imputed GWAS analysis ([S6 Fig.](#) part B; [S7 Table](#)), top SNPs for 1000G imputed data in genes of functional interest ([Table 2](#)) were generally of the same order of magnitude as top genotyped SNPs, including support for the top hit at *BCL9*. Novel findings in the top 100 SNPs for 1000G imputed data included a hit at the previously identified gene for T2D *IGF2BP2* ([S4 Table](#): rs138306797, $P_{\text{imputed}_1000G} = 2.55 \times 10^{-6}$) that had not achieved $P < 10^{-2}$ in the genotyped data, and hits at $P_{\text{imputed}_1000G} < 10^{-6}$ in genes not previously associated genetically or functionally with T2D ([S7 Table](#): *SATB2/TYW5* on chromosome 2, *MTHFD1L/AKAP12* on chromosome 6; *KLF5/KLF12* on chromosome 13).

Genes for obesity or for T2D?

Analysis of T2D and the standardized BMI residual (BMI-longitudinal) was initially performed without any covariate adjustment. Since T2D and BMI are highly correlated and have been shown to have shared genetic influences [68], the question arises as to which phenotype is primarily regulated by the genes of interest identified in this study. We therefore repeated the analyses allowing for BMI and T2D as covariates ([S1 Text](#)). Results demonstrated that (A)— $\log_{10}P$ -values for T2D adjusted for the standardized BMI residual were highly correlated with the original unadjusted analysis for both genotyped and imputed data, and (B) conversely,— $\log_{10}P$ -values for unadjusted BMI and BMI adjusted for T2D were likewise highly correlated. This suggests that genetic effects on T2D as originally calculated are largely independent of any effects attributable to the mean standardized BMI residual. Specifically, the significance levels for genes of functional interest are of the same order of magnitude in both the adjusted ([S1 Text](#)) and the original analysis ([Table 2](#)). Similarly, genetic effects on BMI as originally calculated are largely independent of any effects attributable to T2D, consistent with the regional

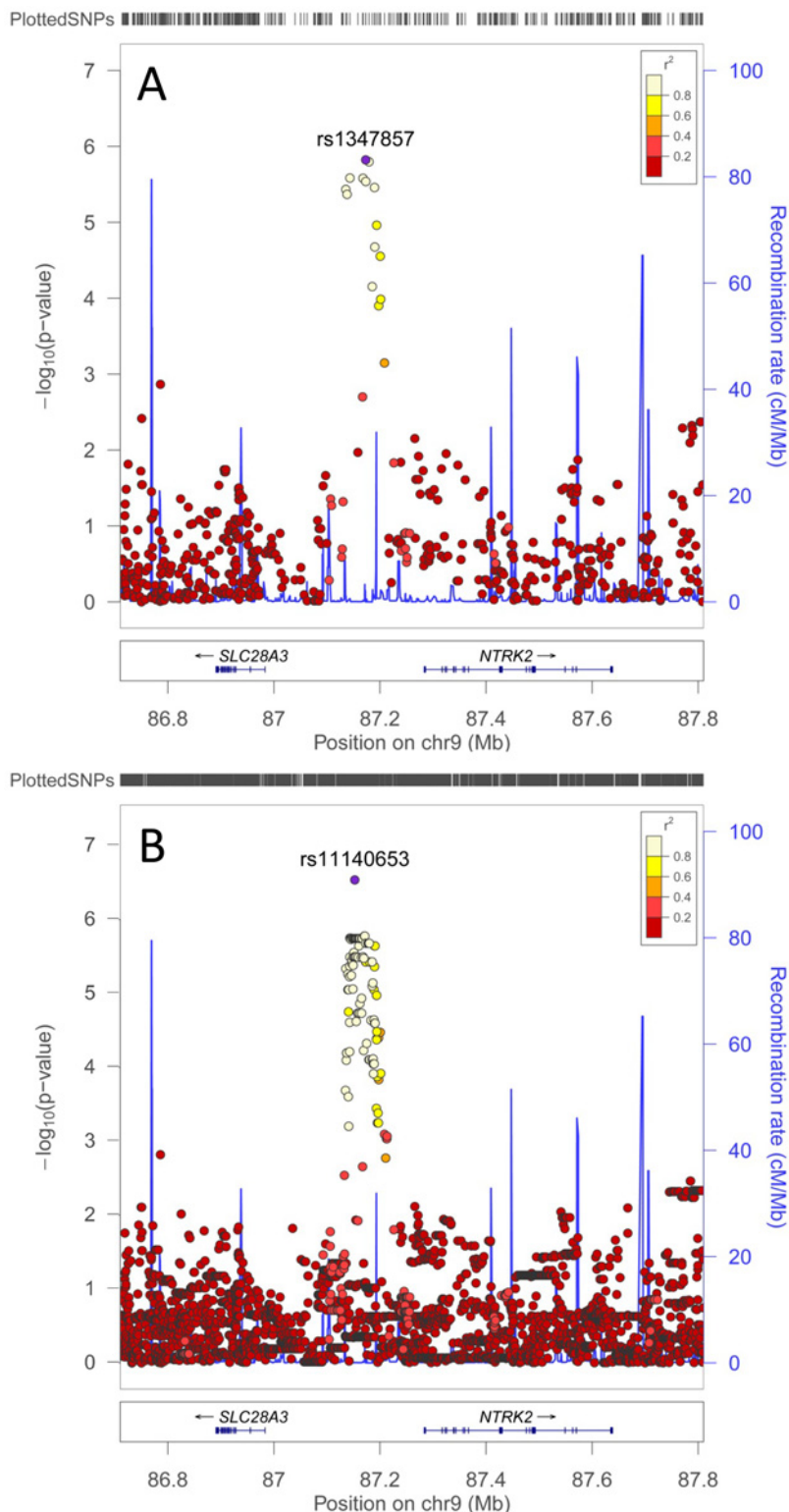


Fig 3. Regional association plots (LocusZoom [35]) of the signal for BMI association in the region *SLC28A3* to *NTRK2* on chromosome 9. (A) is the plot for genotyped data; and (B) is the plot for 1000G imputed data. The $-\log_{10} P$ -values are shown on the upper part of each plot. SNPs are colored (see key) based on their r^2 with the labeled hit SNP (purple), calculated in the 146 unrelated genotyped individuals. The bottom section of each plot shows the genes marked as horizontal lines. The second Y axis is for recombination rate, as shown in blue on the plot.

doi:10.1371/journal.pone.0119333.g003

Table 2. Top GWAS SNP hits in genes of functional interest for T2D, organized by chromosome.

Chromosome	SNP	NCBI37	A1	A2	Genotyped				Imputed 1000G				SNP Location	HGNC*
					effB	se_effB	chi2.1df	P1df	effB	se_effB	chi2.1df	P1df		
1	rs11240074	146996480	A	C	0.21	0.05	20.62	5.59E-06					INTERGENIC	CHD1L/BCL9
	rs11240074	146996480	A	C					0.21	0.05	20.58	5.73E-06	INTERGENIC	CHD1L/BCL9
6	rs6930407	89905239	A	G	0.13	0.03	16.63	4.55E-05					INTRONIC	GABRR1
	rs9451177	89905172	A	G					0.13	0.03	17.37	3.08E-05	INTRONIC	GABRR1
12	rs11063387	4998536	G	C	0.14	0.03	15.50	8.24E-05					INTERGENIC	KCNA6/KCNA1
	rs11063385	4997051	C	T					0.13	0.03	14.97	1.09E-04	INTERGENIC	KCNA6/KCNA1
21	rs8128418	39188732	G	A	0.25	0.06	15.01	1.07E-04					INTRONIC	KCNJ6
	rs113713721	39185153	G	A					0.26	0.06	15.99	6.38E-05	INTRONIC	KCNJ6

Data analyzed in GenABEL using genotyped or 1000G imputed data for all individuals with doctor-diagnosed T2D. Results are for allele-wise tests under an additive model of inheritance. Bold indicates top hit for T2D based on both genotyped and imputed data. Full lists of the top 50 hits for genotyped SNPs, and the top 100 SNPs for imputed data, appear in [S5 Table](#) and [S7 Table](#), respectively.

* Genes separated by forward slash indicate nearest protein coding genes upstream/downstream of the SNP. NCBI37 = bp location on chromosome for NCBI Build 37. A1 = major allele; A2 = minor allele.

doi:10.1371/journal.pone.0119333.t002

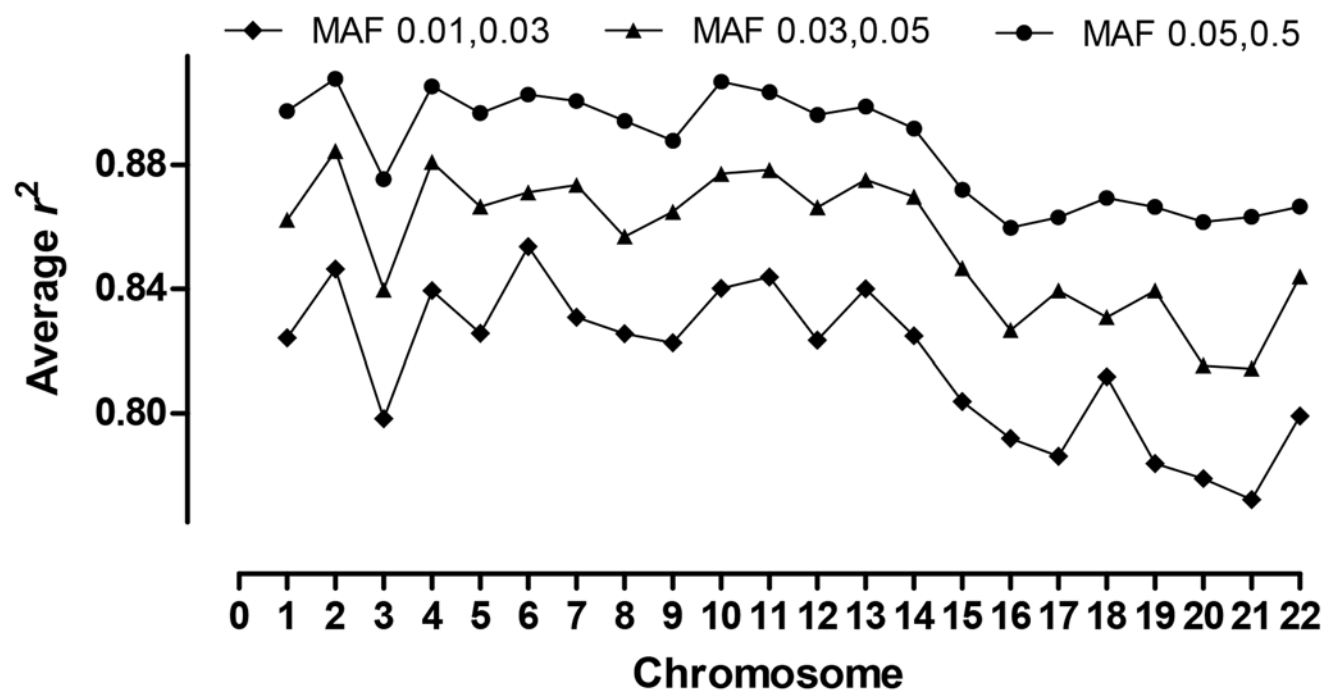


Fig 4. Imputation accuracy for 402 genotyped individuals imputed against the 1000G reference panel. Imputation accuracy is measured as average r^2 across all autosomes for SNPs of different MAFs (see key).

doi:10.1371/journal.pone.0119333.g004

association plots (Fig. 3, S7 Fig.) where no evidence for association is seen in plots comparing T2D results across regions that contained the top hits for BMI. This includes the regions containing *RMB7*, which previously has been related to pancreatic function [53], and *PIK3C2G* which has been associated with T2D in another population [54].

Interrogating the *SLC28A3* to *NTRK2* intergenic region

As with most GWAS for common complex diseases [69], top SNP hits identified here were generally located within introns or in potential regulatory regions upstream or downstream of the gene of interest. In the case of the top hit for BMI (Fig. 3), the peak of association was clearly located upstream of the best functional candidate for BMI, the *NTRK2* gene, with little evidence for linkage disequilibrium between SNPs in this peak of association and SNPs within the coding region of the gene. Conditioning on the top genotyped and top imputed SNPs (S8 Fig.) reduced all other SNP signals to $P > 10^{-3}$, suggesting a single major signal regulating BMI across this region. We therefore interrogated the region of top hits upstream of *NTRK2* to see if we could find evidence for potential regulatory elements. Initially we used SYNPLLOT to plot the location of the top genotyped SNPs ($P < 5 \times 10^{-6}$) in relation to CNS upstream of *NTRK2* (Fig. 5). These 8 genotyped SNPs were all in strong LD with the top SNP ($r^2 > 0.8$; Fig. 3). SNP rs1866439 ($P_{\text{genotyped}} = 1.59 \times 10^{-6}$) was of equivalent significance to the top SNP rs1347857 ($P_{\text{genotyped}} = 1.5 \times 10^{-6}$) and was unique in being the only top SNP located within a CNS peak, with the risk allele causing loss of a NHP1 binding site falling within a half-palindromic estrogen response element (TGAGTagtTG^A/GCC) [70, 71]. *NTRK2* expression has been shown to be regulated by estradiol [72]. Other indications that this might be a regulatory region, as annotated on Fig. 5, included: (i) the presence adjacent to rs1866439 of a CpG-island (50%GC; observed CpG/expected CpG 61%; 212bp in length); (ii) a number of peaks of mono-methylation of lysine 4 of histone H3 (H3K4Me1; frequently found near regulatory regions) as measured by the ENCODE project (see URLs), including around rs1866439; and (iii) the peak of association with the top imputed SNPs (Fig. 3 part B) falling within a non-conserved region (no CNSs) of fragments of long (LINE-1 or L1) interspersed elements or retrotransposons (Fig. 5 and S9 Fig.). S9 Fig. shows the positions of these top SNPs in relation to the L1 elements, with the top imputed SNP (rs11140653) specifically falling within a L1MA3 element (as also annotated on Fig. 5). While full-length L1s are required for retrotransposition [73], regulatory elements that overlap with fragments of L1 sequence have been shown to be involved in gene silencing [74]. Further functional studies will be required to determine which variants and elements are important in regulating *NTRK2* in this Aboriginal population.

Discussion

Results of the discovery GWAS undertaken here provide the first hypothesis-free insights into genetic risk factors for high BMI and T2D in an Australian Aboriginal population. Although estimates suggest the arrival of Aboriginal people in Australia more than 45,000 years ago [66, 67], we found that we were able to (A) genotype with high accuracy using the Illumina 2.5M Duo SNP chip, and (B) impute genotypes with high accuracy based on the reference panel from the 1000 Genomes project [75]. This included imputation accuracy in both the subset of 195 individuals determined by PCA to be of pure Martu ancestry and in the 207 individuals in whom we observed varying degrees of introgression with Caucasian genomes. This imputation accuracy compares favorably with that reported for imputation based on the 1000G reference panel of African and American ancestry groups genotyped on the Illumina Omni2.5M Duo chips [76]. The major limitation in the present study was sample size and power, but as a pioneering study in the area of application of modern genomics to Aboriginal

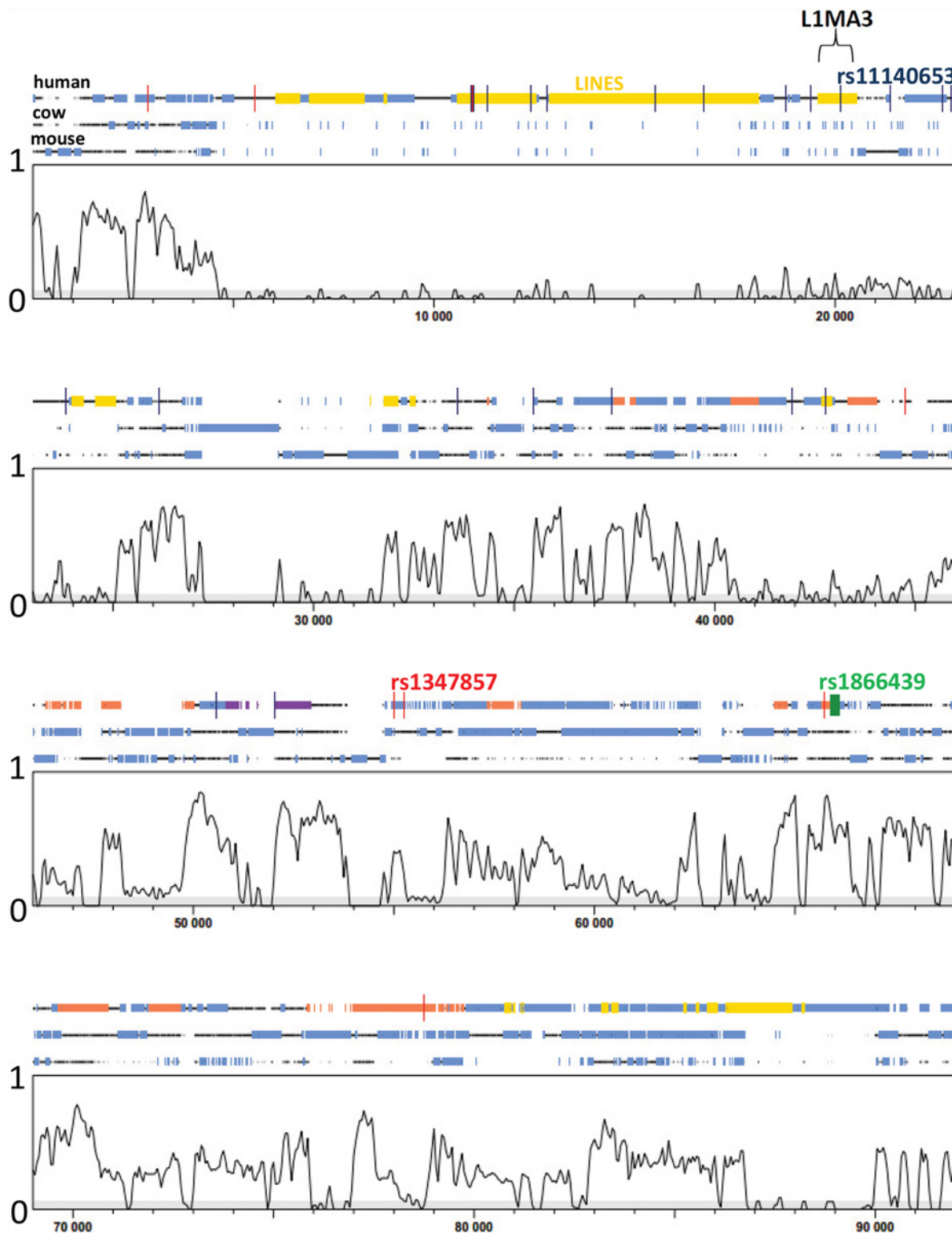


Fig 5. SYNPLLOT [38] for a section of the intergenic region (NCBI Build 37: 87,140,000bp to 87,190,000bp) ~94kb upstream of *NTRK2* on chromosome 9q21.33. Since regulatory elements are usually found with conserved regions of the genome, we interrogated the region of our top hits using this *in silico* analysis to look for conserved regions across multiple species. The multiple alignments of (top to bottom) human, cow and mouse sequences were performed across the complete intergenic region *SLC28A3* to *NTRK2* using LAGAN. The segment of the alignments shown here is the region that contains the top association hits ($P < 5 \times 10^{-6}$). The central plotted curves show the degree of conservation of sequence across all three species, on a vertical

scale 0–1 (= 100%), such that the peaks represent CNS. CNS are defined here as regions with a nucleotide sequence conservation level of ≥ 0.7 , i.e. \geq to the least conserved exon sequence in the two genes (not shown on this plot) *SLC28A3* and *NTRK2* flanking the intergenic region of interest. Blue boxes indicate repetitive sequence in all 3 species. The human sequence is also annotated with positions of: (i) LINE-1 elements (yellow); (ii) the top genotyped SNPs (red vertical bars) including the top SNP rs1086204, and the SNP of interest rs1866439; (iii) the top imputed SNPs (blue vertical bars) including the top SNP rs1140653; (iv) the CpG island-like element identified using CpG Island Seacher (green); and (v) positions of peaks of mono-methylation of lysine 4 of histone H3 (H3K4Me1) as measured in NHEK or NHLF cell lines (mauve) or in H1-hESC human embryonic stem cells (orange) by the ENCODE project, as identified using the UCSC browser (see [S9 Fig.](#)).

doi:10.1371/journal.pone.0119333.g005

health in Australia it is important that results of this study are reported in order to allay concerns about genetic research as applied to health outcomes, to boost confidence in the approach, and stimulate replication studies in other Australian Aboriginal populations. The important practical implication of our demonstration that 1000G imputation could be employed with accuracy in this Australian Aboriginal population is that further cost-effective genome-wide approaches can be applied to understanding pathogenesis of complex diseases in Australian Aboriginal populations.

In our study, modest support ($P < 10^{-2}$) was obtained for a number of genes previously shown to achieve genome wide significance in other populations (see Tables [S2](#) and [S4](#)), but these were generally not represented amongst the top hits observed for our study population. Of note, no SNPs at $P < 10^{-2}$ were observed for *TCF7L2* [[55–59](#)] or *KCNJ11* [[60](#)], genes that have previously been associated with T2D across multiple ethnicities (though not with Pima Indians for *TCF7L2* [[77](#)]). Recent large-scale trans-ancestry meta-analysis shows a significant excess in directional consistency of T2D risk alleles across ancestry groups [[78](#)]. Whilst the power of our study may have limited our ability to replicate top hits found in large scale population-based case-control GWAS and meta-analyses for BMI and T2D, there is increasing support [[79](#), [80](#)] for the possibility that functional variants that are rare in the general population may be enriched through highly shared ancestry, identity-by-descent, and linkage in the kind of extended pedigrees that we have used in our discovery GWAS. Accordingly, our findings here, while requiring further definitive replication in additional Australian Aboriginal populations, highlight interesting novel association signals for BMI and T2D that might provide important clues to disease pathogenesis in this population.

For T2D, the top association in our study population, supported by both genotyped and imputed data, was in the region of *BCL9*. *BCL9* encodes a protein that, along with *TCF7L2*, promotes beta-catenin's transcriptional activity in the WNT-signaling pathway [[81](#)]. WNT-signaling is important in the transcription of proglucagon, which serves as precursor for the incretin hormone glucagon-like peptide 1 that stimulates insulin secretion [[82](#)]. Additional hits of functional interest occurred in *KCNJ6*, *KCNA1*, and *GABRR1*. *KCNJ6* lies adjacent to *KCNJ15* on chromosome 21q22.13. *KCNJ6* is an ATP-sensitive potassium channel that regulates insulin secretion in pancreatic beta cells in response to nutrients [[61](#)]. *KCNJ15* is another member of this family of inwardly-rectifying ATP-sensitive potassium channels that has been shown to be associated with T2D in an Asian population [[83](#)]. *KCNA1* is the voltage-gated potassium channel Kv1.1 that is expressed in pancreatic β -cells and can influence glucose-stimulated insulin release [[62](#)]. It has also been associated with ataxia and epilepsy in humans carrying rare variants [[84](#)], and mutant mice provide a model for ataxia in which Kv1.1 influences gamma-aminobutyric acid (GABA) release in Purkinje cells of the brain [[64](#)]. *GABRR1* is a member of the rho subunit family of GABA A receptors. Previous studies in China have demonstrated association between polymorphisms at *GABRR1* and diabetic cataract [[63](#)]. Recently we also demonstrated that polymorphisms in both GABA-A and GABA-B receptor pathway genes are associated with T2D in an extended family from the United Arab Emirates [[85](#)]. GABA-A receptors act as inhibitory neurotransmitters in the central nervous system. They are

also present in the endocrine part of the pancreas at concentrations comparable to those in the central nervous system, and co-localize with insulin in pancreatic beta cells [86]. Recent work has shown a role for both GABA-A and GABA-B receptors in regulating insulin secretion and glucagon release in pancreatic islet cells from normo-glycaemic and T2D individuals [87].

Amongst our top hits for BMI was *PIK3C2G*, a gene previously associated with T2D and with serum insulin levels in Japan [54]. *PIK3C2G* is a member of a conserved family of intracellular phosphoinositide 3-kinases known to be involved in a large array of cellular functions. Daimon and coworkers [54] related their association to prior observations [88] of expression of *PIK3C2G* in the pancreas. However, recent GWAS have observed associations between polymorphisms at *PIK3C2G* and hyperlipidemia and myocardial infarction [89].

A second hit for BMI in our population was *CNTNAP2* that encodes contactin associated protein-like 2 which localizes to the juxtaparanodal region of the nodes of Ranvier in myelinated axons, where it is required for proper localization of the potassium voltage-gated channel KCNA1 [90]. Mutations in *Cntnap2* in mice impair Kv1.1 localization, and can be obesity-promoting or obesity-resistant in diet-induced obesity depending on genetic background [91]. Polymorphism at *CNTNAP2* was among the strongest gene x environment interactions among African-Americans with BMI as outcome, specifically interacting with dietary energy intake [92]. *CNTNAP2* is amongst genes that lie within regions of *de novo* duplications and deletions recently linked to syndromic obesity in children [93].

Perhaps the most convincing association observed in our study was between BMI and SNPs that lie in the intergenic region between *SLC28A3* and *NTRK2* on chromosome 9q21.33. *SLC28A3* is a concentrative nucleoside transporter with broad specificity for pyrimidine and purine nucleosides [94], and is therefore not a strong candidate genetic risk factor for BMI. On the other hand, *NTRK2* (also known as *TRKB* or tyrosine kinase receptor B) is the receptor for BDNF and regulates energy balance downstream of MC4R. Associations between BMI and SNPs at both *MC4R* [95–97] and *BDNF* [98, 99] have achieved genome-wide significance in multiple independent studies, but association between *NTRK2* and BMI has only previously been observed ($P = 1.04 \times 10^{-6}$) in a gene-centric multi-ethnic meta-analysis of 108,912 individuals genotyped on the ITMAT-Broad-Candidate Gene Association Resource (CARE) array containing 49,320 SNPs across 2100 metabolic and cardiovascular-related loci [100]. Mouse *Bdnf* mutants that express decreased amounts of Ntrk2 show hyperphagia and maturity-onset obesity [101], while risk variants in human *BDNF* were significantly associated with more food servings in a study of obesity susceptibility loci and dietary intake [102]. Deficiency in MC4R signaling reduces expression of BDNF in ventromedial hypothalamic nuclei, indicating that BDNF and its receptor *NTRK2* are downstream components in the MC4R-mediated control of energy balance. Overall, our data are consistent with meta-analysis [96] of large-scale GWAS of BMI that, along with data from rare monogenic forms of obesity [103], highlight a neuronal influence on body weight regulation.

In summary—we report here data for the first GWAS of complex disease in an Australian Aboriginal population. Whilst the top hits for BMI and T2D are in novel genes not yet reported for other GWAS, they occur in genes that belong to key pathways strongly supported by previous GWAS in other ethnicities. This means that current international efforts [82, 104, 105] to target these pathways will be relevant and translatable to the Australian Aboriginal population, and the genes identified here may indeed provide novel targets in this endeavour.

Online Resources

The URLs for web resources presented herein are as follows:

R version 2.15.0, <http://www.R-project.org/>

The `mhtplot()` function of 'gap' for R, <http://www.mrc-epid.cam.ac.uk/Personal/jinghua.zhao/r-progs.htm>

UCSC genome-browser, <http://ucscbrowser.genenetwork.org/cgi-bin/hgGateway>

The Encode Project, <http://www.genome.gov/encode/>

Smit AFA, Hubley R, Green P (1996–2010) RepeatMasker Open-3.0. <http://www.repeatmasker.org>

Supporting Information

S1 Fig. Radial plot showing hierarchical clustering of estimated pairwise identity-by-descent allele-sharing for the 402 genotyped individuals used in this study. The genomic kinship matrix was first calculated in GenABEL v1.7–6, and converted to a distance matrix and hierarchical cluster analysis using single linkage on the dissimilarities. The ape package was used to produce radial tree plot.

(PDF)

S2 Fig. Principal component (PC) analysis (PCA) plots showing population substructure in the study population. A subset of 70,420 genotyped SNPs with pairwise linkage disequilibrium (LD; r^2) ≤ 0.3 and MAF > 0.01 was used in PCA (SMARTPCA within EIGENSOFT) to look at population substructure across the 402 genotyped family members. Plots (A) PC1 x PC2, (B) PC1xPC3, and (C) PC2xPC3 show individuals, color coded by age (see key). (D) shows the PC1 x PC2 plot in which individuals are color coded (see key) according to their T2D status. PCA plots (E) PC1 x PC2, (F) PC1xPC3, and (G) PC2xPC3 show individuals color coded by ancestry (see key). Note that the Martu clusters comprise 73 individuals aged < 20 , 25 individuals aged 20–30, 57 individuals aged 30–50, and 40 individuals aged > 50 . Hence there is no evidence for disproportionate representation in any age class. Reference HapMap populations are not included at the specific request of the Board of the Aboriginal Health Service.

(PDF)

S3 Fig. Demonstrates the strongly correlated GWAS results (genotyped data) for mean BMI compared with BMI longitudinal for (A) FASTA GenABEL, and (B) Fast-LMM analyses.

(PDF)

S4 Fig. Demonstrates the strongly correlated GWAS results (genotyped data) for FASTA GenABEL versus Fast-LMM for (A) mean BMI, and (B) BMI longitudinal.

(PDF)

S5 Fig. Provides Q-Q plots and associated lambda values for GWAS analyses undertaken for (A) FASTA GenABEL BMI longitudinal, (B) FASTA GenABEL mean BMI, (C) Fast-LMM BMI longitudinal, (D) Fast-LMM mean BMI.

(PDF)

S6 Fig. Manhattan plots of genome-wide association results for T2D undertaken using FASTA in GenABEL. (A) results for genotyped data; and (B) results for 1000G imputed data. SNPs in red show the region of the top association in this discovery GWAS.

(PDF)

S7 Fig. Regional association plots (Locuszoom) of the imputed SNP signals for BMI (upper graph) and T2D (lower graph) in specific regions. (A) *SLC28A3* to *NTRK2* on Chromosome 9; (B) *CNTNAP2* on Chromosome 7; (C) *RBM7* on Chromosome 4; and (D) *PIK3C2G* on Chromosome 12. In each plot the $-\log_{10}$ *P*-values are shown on the upper section, with SNPs

colored (see key) based on their r^2 with the labeled top hit SNP (purple), calculated in the 146 unrelated genotyped individuals. Red arrows highlight the position of the BMI hit on the T2D plot. The bottom section of each plot shows the genes marked as horizontal lines.
(PDF)

S8 Fig. Regional association plots (Locuszoom) of the signals for (A) BMI genotyped and (B) BMI imputed in the region *SLC28A3* to *NTRK2* on Chromosome 9 before (upper graph) and after (lower graph) conditioning on the top SNPs rs1347857 and rs11140653, respectively. Analysis undertaken using an additive model in FaST-LMM.
(PDF)

S9 Fig. Plot generated in the UCSC genome-browser for a section of the intergenic region (NCBI Build 37: 87,140,000bp to 87,190,000bp) ~94kb upstream of *NTRK2* on Chromosome 9q21.33. The plot shows positions of the top ($P < 5 \times 10^{-6}$) genotyped (in red vertical lines) and 1000G imputed (in blue vertical lines) SNPs, with the top imputed SNP (rs1140653), the top genotyped SNP (rs1347857) and the SNP of functional interest discussed in the main text (rs1866439) shown again on separate rows. These SNPs are shown relative to: (i) the CpG island-like motif identified using CpG Island Searcher; (ii) the various fragments of LINE-1 (L1) elements (with the L1MA3 element in which the top imputed SNP is located shown in blue); and (iii) peaks of histone methylation (H3K4Me1, H3K4Me3) or acetylation (H3K27Ac) measured in 7 cell lines from the ENCODE project.
(PDF)

S1 Table. Characteristics of the family-based Aboriginal study sample. The 391 post-QC genotyped individuals used in the GWAS belonged to a small number of interrelated extended pedigrees, as depicted in the radial plot of pairwise identity-by-descent allele-sharing presented in [S1 Fig](#).
Characteristics of the 391 genotyped individuals used in the GWAS are provided here.
(PDF)

S2 Table. SNP associations at $P < 0.01$ observed in the WA Aboriginal study population for genes previously reported* to achieve $P < 5 \times 10^{-8}$ (or $P < 10^{-5}$ shaded grey) for association with BMI in other populations. Bold indicates a SNP hits observed in the imputed but not the genotyped data, including at *DICER* ($P < 10^{-6}$).
(PDF)

S3 Table. Top GWAS autosomal SNP hits for BMI, organised by chromosome. Results for the top 50 hits for “BMI longitudinal” GWAS analysis in GenABEL using each individual BMI reading as a separate observation, modelling the correlation between readings via the estimated kinship and using the genomic control deflation factor to avoid inflation of the overall distribution of test statistics. Results are for allele-wise tests under an additive model of inheritance. Bold indicates SNP associations of functional interest presented in main [Table 1](#).
(PDF)

S4 Table. SNP associations observed at nominal $P < 0.01$ in the WA Aboriginal study population for genes previously reported to achieve $P < 5 \times 10^{-8}$ for association with T2D in other populations. Bold indicates SNP hits observed in the imputed but not the genotyped data, including at *IGF2BP2* ($P < 10^{-5}$).
(PDF)

S5 Table. Top 50 GWAS SNP hits for T2D, organised by chromosome. Results are for allele-wise tests under an additive model of inheritance. Bold indicates SNP associations of functional

interest presented in main [Table 2](#).
(PDF)

S6 Table. Top GWAS imputed SNP hits for BMI, organised by chromosome. Results for the top 100 hits for “BMI longitudinal” GWAS analysis in GenABEL using each individual BMI reading as a separate observation, modelling the correlation between readings via the estimated kinship and using the genomic control deflation factor to avoid inflation of the overall distribution of test statistics. Results are for allele-wise tests under an additive model of inheritance. Bold indicates 3 top SNP associations for imputed data ($P < 10^{-6}$) coincide with two top genes of functional interest (*NTRK2*, *PIK3C2G*) presented in main [Table 1](#), as well as a hit near *DICER* a previously observed GWAS hit for BMI (see [S2 Table](#)).
(PDF)

S7 Table. Top GWAS imputed SNP hits for T2D, organised by chromosome. Results are for the top 100 hits for T2D GWAS analysis in GenABEL for allele-wise tests under an additive model. Bold indicates 3 top SNP associations for imputed data ($P < 10^{-6}$) not observed in the genotyped data. Bold plus grey shading indicates hits that coincide with top gene of functional interest (*BCL9*) presented in main [Table 2](#), as well as a hit near *IGF2BP2* a previously observed GWAS hit for T2D (see [S4 Table](#)).
(PDF)

S1 Text. Supplement on Conditional Analyses.
(PDF)

Acknowledgments

We gratefully acknowledge the tremendous contribution made by the Aboriginal community, the Board and the staff of the local Aboriginal Health Service (AHS) where our study was based, and the support of local schools in the area. Without this support the study would not have been possible, nor would it have been possible to engage the community in a “Feedback to Community” project that resulted in production of the animation “The Goanna and the Journey of the Genes” available in Martu (<https://www.youtube.com/watch?v=05uNhejT7Ik>) and English (<https://www.youtube.com/watch?v=xrGRfrRGBLE&noredirect=1>) language versions. We thank Steve Aiton (animation facilitator) and Ruth Koedyk (artist) for their assistance with the feedback to community projects. We also acknowledge the generous in-kind support provided by the AHS for travel and accommodation to allow the field collection of samples used in the study, and the generosity of the Board of the AHS in allowing access to Communicare records through our Memorandum of Understanding. Our thanks also go to members of WAAHEC for their helpful contributions and insights as we prepared ethical approvals for this study, and to Associate Professor Ted Wilkes (Curtin University), Mr Glenn Pearson (Manager, Aboriginal Health Research, Telethon Kids Institute), and Dr Emma Kowal (University of Melbourne) for sharing that journey with us.

Author Contributions

Conceived and designed the experiments: HJC ED SEJ JMB. Performed the experiments: GS ESHS SEJ JMB. Analyzed the data: DA HJC MF RWF GS JMB. Wrote the paper: DA HJC MF SEJ JMB. Clinical diagnosis, clinical care, and assistance with recruitment: SJM TM.

References

1. Donnelly P. Genome-sequencing anniversary. Making sense of the data. *Science*. 2011; 331(6020):1024–5. doi: [10.1126/science.1204089](https://doi.org/10.1126/science.1204089) PMID: [21350161](https://pubmed.ncbi.nlm.nih.gov/21350161/)
2. Fesinmeyer MD, North KE, Lim U, Buzkova P, Crawford DC, Haessler J, et al. Effects of smoking on the genetic risk of obesity: the population architecture using genomics and epidemiology study. *BMC Med Genet*. 2013; 14:6. doi: [10.1186/1471-2350-14-6](https://doi.org/10.1186/1471-2350-14-6) PMID: [23311614](https://pubmed.ncbi.nlm.nih.gov/23311614/)
3. McCarthy MI. Genomics, type 2 diabetes, and obesity. *N Engl J Med*. 2010; 363(24):2339–50. doi: [10.1056/NEJMr0906948](https://doi.org/10.1056/NEJMr0906948) PMID: [21142536](https://pubmed.ncbi.nlm.nih.gov/21142536/)
4. Morris AP, Voight BF, Teslovich TM, Ferreira T, Segre AV, Steinthorsdottir V, et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat Genet*. 2012; 44(9):981–90. doi: [ng.2383](https://doi.org/10.1038/ng.2383) PMID: [22885922](https://pubmed.ncbi.nlm.nih.gov/22885922/)
5. Scott RA, Lagou V, Welch RP, Wheeler E, Montasser ME, Luan J, et al. Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. *Nat Genet*. 2012; 44(9):991–1005. doi: [ng.2385](https://doi.org/10.1038/ng.2385) PMID: [22885924](https://pubmed.ncbi.nlm.nih.gov/22885924/)
6. Xi B, Takeuchi F, Meirhaeghe A, Kato N, Chambers JC, Morris AP, et al. Associations of genetic variants in/near body mass index-associated genes with type 2 diabetes: a systematic meta-analysis. *Clin Endocrinol (Oxf)*. 2014; 81(5):702–10. doi: [10.1111/cen.12428](https://doi.org/10.1111/cen.12428) PMID: [24528214](https://pubmed.ncbi.nlm.nih.gov/24528214/)
7. Gillett G, McKergow F. Genes, ownership, and indigenous reality. *Social science & medicine*. 2007; 65(10):2093–104. doi: [10.1016/j.socscimed.2007.06.024](https://doi.org/10.1016/j.socscimed.2007.06.024)
8. McInnes RR. 2010 Presidential Address: Culture: the silent language geneticists must learn—genetic research with indigenous populations. *Am J Hum Genet*. 2011; 88(3):254–61. PMID: [21516613](https://pubmed.ncbi.nlm.nih.gov/21516613/)
9. Kowal E, Anderson I. Difficult conversations: talking about Indigenous genetic research in Australia. In: Berthier S, Tolazzi S, Whittick S, editors. *Biomapping—Indigenous Identities*. Amsterdam: Rodopi; 2012.
10. Kowal E, Pearson G, Peacock CS, Jamieson SE, Blackwell JM. Genetic research and aboriginal and Torres Strait Islander Australians. *J Bioeth Inq*. 2012; 9(4):419–32. doi: [10.1007/s11673-012-9391-x](https://doi.org/10.1007/s11673-012-9391-x) PMID: [23188401](https://pubmed.ncbi.nlm.nih.gov/23188401/)
11. Kowal EE. Genetic research in Indigenous health: significant progress, substantial challenges. *Comment. Med J Aust*. 2012; 197(7):384. doi: [10.5694/mja.12.11113](https://doi.org/10.5694/mja.12.11113) PMID: [23025733](https://pubmed.ncbi.nlm.nih.gov/23025733/)
12. Kowal EE. Genetic research in Indigenous health: significant progress, substantial challenges. *Med J Aust*. 2012; 197(1):19–20. doi: [10.5694/mja.12.10531](https://doi.org/10.5694/mja.12.10531) PMID: [22762219](https://pubmed.ncbi.nlm.nih.gov/22762219/)
13. Dodson M, Williamson R. Indigenous peoples and the morality of the Human Genome Diversity Project. *J Med Ethics*. 1999; 25(2):204–8. PMID: [10226929](https://pubmed.ncbi.nlm.nih.gov/10226929/)
14. Kowal E, Anderson I. Genetic research in Aboriginal and Torres Strait Islander communities: Continuing the conversation: http://www.lowitja.org.au/files/docs/Genetics_report2012.pdf; 2012. Available from: http://www.lowitja.org.au/files/docs/Genetics_report2012.pdf.
15. Rotimi CN, Jorde LB. Ancestry and disease in the age of genomic medicine. *N Engl J Med*. 2010; 363(16):1551–8. doi: [10.1056/NEJMr0911564](https://doi.org/10.1056/NEJMr0911564) PMID: [20942671](https://pubmed.ncbi.nlm.nih.gov/20942671/)
16. Daniel M, Rowley KG, McDermott R, O'Dea K. Diabetes and impaired glucose tolerance in Aboriginal Australians: prevalence and risk. *Diabetes Res Clin Pract*. 2002; 57(1):23–33. PMID: [12007727](https://pubmed.ncbi.nlm.nih.gov/12007727/)
17. Tonkinson R. *The Mardu Aborigines, living the dream in Australia's desert*. New York: Holt, Reinhart & Winston; 1991.
18. Davenport S, Johnson PC, Yuwali. *Cleared Out: First Contract in the Western Desert*. Aboriginal Studies Press; 2005.
19. Alberti KG, Zimmet PZ. Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: diagnosis and classification of diabetes mellitus provisional report of a WHO consultation. *Diabet Med*. 1998; 15(7):539–53. doi: [10.1002/\(SICI\)1096-9136\(199807\)15:7<539::AID-DIA668>3.0.CO;2-S](https://doi.org/10.1002/(SICI)1096-9136(199807)15:7<539::AID-DIA668>3.0.CO;2-S) PMID: [9686693](https://pubmed.ncbi.nlm.nih.gov/9686693/)
20. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. *PLoS Genet*. 2006; 2(12):e190. doi: [10.1371/journal.pgen.0020190](https://doi.org/10.1371/journal.pgen.0020190) PMID: [17194218](https://pubmed.ncbi.nlm.nih.gov/17194218/)
21. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*. 2006; 38(8):904–9. doi: [ng1847](https://doi.org/10.1038/ng1847) PMID: [16862161](https://pubmed.ncbi.nlm.nih.gov/16862161/)
22. Delaneau O, Marchini J, Zagury JF. A linear complexity phasing method for thousands of genomes. *Nat Methods*. 2012; 9(2):179–81. doi: [10.1038/nmeth.1785](https://doi.org/10.1038/nmeth.1785) PMID: [217851785](https://pubmed.ncbi.nlm.nih.gov/217851785/)
23. Howie B, Marchini J, Stephens M. Genotype imputation with thousands of genomes. *G3 (Bethesda)*. 2011; 1(6):457–70. doi: [10.1534/g3.111.001198](https://doi.org/10.1534/g3.111.001198) PMID: [21198001198](https://pubmed.ncbi.nlm.nih.gov/21198001198/)

24. Marchini J, Howie B. Genotype imputation for genome-wide association studies. *Nat Rev Genet*. 2010; 11(7):499–511. doi: [10.1038/nrg2796](https://doi.org/10.1038/nrg2796) [pii] PMID: [20517342](https://pubmed.ncbi.nlm.nih.gov/20517342/)
25. Lin P, Hartz SM, Zhang Z, Saccone SF, Wang J, Tischfield JA, et al. A new statistic to evaluate imputation reliability. *PLoS ONE*. 2010; 5(3):e9697. doi: [10.1371/journal.pone.0009697](https://doi.org/10.1371/journal.pone.0009697) PMID: [20300623](https://pubmed.ncbi.nlm.nih.gov/20300623/)
26. Liu EY, Buyske S, Aragaki AK, Peters U, Boerwinkle E, Carlson C, et al. Genotype imputation of Meta-chip SNPs using a study-specific reference panel of ~4,000 haplotypes in African Americans from the Women's Health Initiative. *Genet Epidemiol*. 2012; 36(2):107–17. doi: [10.1002/gepi.21603](https://doi.org/10.1002/gepi.21603) PMID: [22851474](https://pubmed.ncbi.nlm.nih.gov/22851474/)
27. Abecasis GR, Cardon LR, Cookson WO. A general test of association for quantitative traits in nuclear families. *Am J Hum Genet*. 2000; 66:279–92. PMID: [10631157](https://pubmed.ncbi.nlm.nih.gov/10631157/)
28. Aulchenko YS, Ripke S, Isaacs A, van Duijn CM. GenABEL: an R library for genome-wide association analysis. *Bioinformatics*. 2007; 23(10):1294–6. doi: [btm108](https://doi.org/10.1093/bioinformatics/btm108) [pii] [10.1093/bioinformatics/btm108](https://doi.org/10.1093/bioinformatics/btm108) PMID: [17384015](https://pubmed.ncbi.nlm.nih.gov/17384015/)
29. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet*. 2011; 88(1):76–82. doi: [10.1016/j.ajhg.2010.11.011](https://doi.org/10.1016/j.ajhg.2010.11.011) S0002-9297(10)00598-7 [pii] PMID: [21167468](https://pubmed.ncbi.nlm.nih.gov/21167468/)
30. Paradis E, Claude J, Strimmer K. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics*. 2004; 20(2):289–90. PMID: [14734327](https://pubmed.ncbi.nlm.nih.gov/14734327/)
31. Chen WM, Abecasis GR. Family-based association tests for genomewide association scans. *Am J Hum Genet*. 2007; 81(5):913–26. doi: [S0002-9297\(07\)63869-5](https://doi.org/10.1086/521580) [pii] [10.1086/521580](https://doi.org/10.1086/521580) PMID: [17924335](https://pubmed.ncbi.nlm.nih.gov/17924335/)
32. Lippert C, Listgarten J, Liu Y, Kadie CM, Davidson RI, Heckerman D. FaST linear mixed models for genome-wide association studies. *Nat Methods*. 2011; 8(10):833–5. doi: [nmeth.1681](https://doi.org/10.1038/nmeth.1681) [pii] [10.1038/nmeth.1681](https://doi.org/10.1038/nmeth.1681) PMID: [21892150](https://pubmed.ncbi.nlm.nih.gov/21892150/)
33. Kang HM, Sul JH, Service SK, Zaitlen NA, Kong SY, Freimer NB, et al. Variance component model to account for sample structure in genome-wide association studies. *Nat Genet*. 2010; 42(4):348–54. doi: [ng.548](https://doi.org/10.1038/ng.548) [pii] [10.1038/ng.548](https://doi.org/10.1038/ng.548) PMID: [20208533](https://pubmed.ncbi.nlm.nih.gov/20208533/)
34. Gauderman WJ. Sample size requirements for association studies of gene-gene interaction. *Am J Epidemiol*. 2002; 155(5):478–84. PMID: [11867360](https://pubmed.ncbi.nlm.nih.gov/11867360/)
35. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Glied TP, et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics*. 2010; 26(18):2336–7. doi: [btq419](https://doi.org/10.1093/bioinformatics/btq419) [pii] [10.1093/bioinformatics/btq419](https://doi.org/10.1093/bioinformatics/btq419) PMID: [20634204](https://pubmed.ncbi.nlm.nih.gov/20634204/)
36. Brudno M, Do CB, Cooper GM, Kim MF, Davydov E, Green ED, et al. LAGAN and Multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA. *Genome Res*. 2003; 13(4):721–31. doi: [10.1101/gr.926603](https://doi.org/10.1101/gr.926603) GR-9266R [pii] PMID: [12654723](https://pubmed.ncbi.nlm.nih.gov/12654723/)
37. Brudno M, Chapman M, Gottgens B, Batzoglu S, Morgenstern B. Fast and sensitive multiple alignment of large genomic sequences. *BMC Bioinformatics*. 2003; 4:66. doi: [10.1186/1471-2105-4-66](https://doi.org/10.1186/1471-2105-4-66) [pii] PMID: [14693042](https://pubmed.ncbi.nlm.nih.gov/14693042/)
38. Gottgens B, Gilbert JG, Barton LM, Grafham D, Rogers J, Bentley DR, et al. Long-range comparison of human and mouse SCL loci: localized regions of sensitivity to restriction endonucleases correspond precisely with peaks of conserved noncoding sequences. *Genome Res*. 2001; 11(1):87–97. PMID: [11156618](https://pubmed.ncbi.nlm.nih.gov/11156618/)
39. Messeguer X, Escudero R, Farre D, Nunez O, Martinez J, Alba MM. PROMO: detection of known transcription regulatory elements using species-tailored searches. *Bioinformatics*. 2002; 18(2):333–4. PMID: [11847087](https://pubmed.ncbi.nlm.nih.gov/11847087/)
40. Grabe N. AliBaba2: context specific identification of transcription factor binding sites. *In Silico Biol*. 2002; 2(1):S1–15. PMID: [11808873](https://pubmed.ncbi.nlm.nih.gov/11808873/)
41. Cartharius K, Frech K, Grote K, Klocke B, Haltmeier M, Klingenhoff A, et al. MatInspector and beyond: promoter analysis based on transcription factor binding sites. *Bioinformatics*. 2005; 21(13):2933–42. PMID: [15860560](https://pubmed.ncbi.nlm.nih.gov/15860560/)
42. Takai D, Jones PA. Comprehensive analysis of CpG islands in human chromosomes 21 and 22. *Proc Natl Acad Sci U S A*. 2002; 99(6):3740–5. PMID: [11891299](https://pubmed.ncbi.nlm.nih.gov/11891299/)
43. Cole TJ, Donaldson MD, Ben-Shlomo Y. SITAR—a useful instrument for growth curve analysis. *Int J Epidemiol*. 2010; 39(6):1558–66. doi: [10.1093/ije/dyq115](https://doi.org/10.1093/ije/dyq115) PMID: [20647267](https://pubmed.ncbi.nlm.nih.gov/20647267/)
44. Kuczmarski RJ, Flegal KM. Criteria for definition of overweight in transition: background and recommendations for the United States. *Am J Clin Nutr*. 2000; 72(5):1074–81. PMID: [11063431](https://pubmed.ncbi.nlm.nih.gov/11063431/)
45. Haslam DW, James WP. Obesity. *The Lancet*. 2005; 366:1197–209. PMID: [16198769](https://pubmed.ncbi.nlm.nih.gov/16198769/)

46. van Dongen J, Willemsen G, Chen WM, de Geus EJ, Boomsma DI. Heritability of metabolic syndrome traits in a large population-based sample. *J Lipid Res*. 2013; 54(10):2914–23. doi: [10.1194/jlr.P041673](https://doi.org/10.1194/jlr.P041673) PMID: [23918046](https://pubmed.ncbi.nlm.nih.gov/23918046/)
47. Ning F, Silventoinen K, Pang ZC, Kaprio J, Wang SJ, Zhang D, et al. Genetic and environmental correlations between body mass index and waist circumference in China: the Qingdao adolescent twin study. *Behav Genet*. 2013; 43(4):340–7. doi: [10.1007/s10519-013-9597-7](https://doi.org/10.1007/s10519-013-9597-7) PMID: [23756614](https://pubmed.ncbi.nlm.nih.gov/23756614/)
48. Johnson W, Choh AC, Lee M, Towne B, Czerwinski SA, Demerath EW. Characterization of the infant BMI peak: sex differences, birth year cohort effects, association with concurrent adiposity, and heritability. *Am J Hum Biol*. 2013; 25(3):378–88. doi: [10.1002/ajhb.22385](https://doi.org/10.1002/ajhb.22385) PMID: [23606227](https://pubmed.ncbi.nlm.nih.gov/23606227/)
49. Devlin B, Roeder K. Genomic control for association studies. *Biometrics*. 1999; 55(4):997–1004. PMID: [11315092](https://pubmed.ncbi.nlm.nih.gov/11315092/)
50. Eu-Ahsunthornwattana J, Miller EN, Fakiola M, Wellcome Trust Case Control C, Jeronimo SM, Blackwell JM, et al. Comparison of methods to account for relatedness in genome-wide association studies with family-based data. *PLoS Genet*. 2014; 10(7):e1004445. doi: [10.1371/journal.pgen.1004445](https://doi.org/10.1371/journal.pgen.1004445) PMID: [25033443](https://pubmed.ncbi.nlm.nih.gov/25033443/)
51. Dudbridge F, Gusnanto A. Estimation of significance thresholds for genomewide association scans. *Genet Epidemiol*. 2008; 32:227–34. doi: [10.1002/gepi.20297](https://doi.org/10.1002/gepi.20297) PMID: [18300295](https://pubmed.ncbi.nlm.nih.gov/18300295/)
52. Hindorf LA, Junkins HA, Mehta JP, Manolio TA. A Catalog of Published Genome-Wide Association Studies. Available at: www.genome.gov/26525384. Accessed [25 October 2014]. 2009.
53. Harb G, Vasavada RC, Cobrinik D, Stewart AF. The retinoblastoma protein and its homolog p130 regulate the G1/S transition in pancreatic beta-cells. *Diabetes*. 2009; 58(8):1852–62. doi: [10.2337/db08-0759](https://doi.org/10.2337/db08-0759) PMID: [19509021](https://pubmed.ncbi.nlm.nih.gov/19509021/)
54. Daimon M, Sato H, Oizumi T, Toriyama S, Saito T, Karasawa S, et al. Association of the PIK3C2G gene polymorphisms with type 2 DM in a Japanese population. *Biochem Biophys Res Commun*. 2008; 365(3):466–71. doi: [10.1016/j.bbrc.2007.10.180](https://doi.org/10.1016/j.bbrc.2007.10.180) PMID: [17991425](https://pubmed.ncbi.nlm.nih.gov/17991425/)
55. Alsmadi O, Al-Rubeaan K, Mohamed G, Alkayal F, Al-Saud H, Al-Saud NA, et al. Weak or no association of TCF7L2 variants with Type 2 diabetes risk in an Arab population. *BMC Med Genet*. 2008; 9:72. doi: [10.1186/1471-2350-9-72](https://doi.org/10.1186/1471-2350-9-72) PMID: [18655717](https://pubmed.ncbi.nlm.nih.gov/18655717/)
56. Chandak GR, Janipalli CS, Bhaskar S, Kulkarni SR, Mohankrishna P, Hattersley AT, et al. Common variants in the TCF7L2 gene are strongly associated with type 2 diabetes mellitus in the Indian population. *Diabetologia*. 2007; 50(1):63–7. doi: [10.1007/s00125-006-0502-2](https://doi.org/10.1007/s00125-006-0502-2) PMID: [17093941](https://pubmed.ncbi.nlm.nih.gov/17093941/)
57. Ereqat S, Nasereddin A, Cauchi S, Azmi K, Abdeen Z, Amin R. Association of a common variant in TCF7L2 gene with type 2 diabetes mellitus in the Palestinian population. *Acta Diabetol*. 2010; 47 (Suppl 1):195–8. doi: [10.1007/s00592-009-0161-0](https://doi.org/10.1007/s00592-009-0161-0) PMID: [19885641](https://pubmed.ncbi.nlm.nih.gov/19885641/)
58. Lyssenko V, Lupi R, Marchetti P, Del Guerra S, Orho-Melander M, Almgren P, et al. Mechanisms by which common variants in the TCF7L2 gene increase risk of type 2 diabetes. *J Clin Invest*. 2007; 117 (8):2155–63. doi: [10.1172/JCI30706](https://doi.org/10.1172/JCI30706) PMID: [17671651](https://pubmed.ncbi.nlm.nih.gov/17671651/)
59. Saadi H, Nagelkerke N, Carruthers SG, Benedict S, Abdulkhalek S, Reed R, et al. Association of TCF7L2 polymorphism with diabetes mellitus, metabolic syndrome, and markers of beta cell function and insulin resistance in a population-based sample of Emirati subjects. *Diabetes Res Clin Pract*. 2008; 80(3):392–8. doi: [10.1016/j.diabres.2008.01.008](https://doi.org/10.1016/j.diabres.2008.01.008) PMID: [18282631](https://pubmed.ncbi.nlm.nih.gov/18282631/)
60. Qin LJ, Lv Y, Huang QY. Meta-analysis of association of common variants in the KCNJ11-ABCC8 region with type 2 diabetes. *Genet Mol Res*. 2013; 12(3):2990–3002. doi: [10.4238/2013.August.20.1](https://doi.org/10.4238/2013.August.20.1) PMID: [24065655](https://pubmed.ncbi.nlm.nih.gov/24065655/)
61. Sakura H, Bond C, Warren-Perry M, Horsley S, Kearney L, Tucker S, et al. Characterization and variation of a human inwardly-rectifying-K-channel gene (KCNJ6): a putative ATP-sensitive K-channel subunit. *FEBS Lett*. 1995; 367(2):193–7. PMID: [7796919](https://pubmed.ncbi.nlm.nih.gov/7796919/)
62. Ma Z, Lavebratt C, Almgren M, Portwood N, Forsberg LE, Branstrom R, et al. Evidence for presence and functional effects of Kv1.1 channels in beta-cells: general survey and results from mceph/mceph mice. *PLoS ONE*. 2011; 6(4):e18213. doi: [10.1371/journal.pone.0018213](https://doi.org/10.1371/journal.pone.0018213) PMID: [21483673](https://pubmed.ncbi.nlm.nih.gov/21483673/)
63. Lin HJ, Huang YC, Lin JM, Liao WL, Wu JY, Chen CH, et al. Novel susceptibility genes associated with diabetic cataract in a Taiwanese population. *Ophthalmic Genet*. 2013; 34(1–2):35–42. doi: [10.3109/13816810.2012.736590](https://doi.org/10.3109/13816810.2012.736590)
64. Herson PS, Virk M, Rustay NR, Bond CT, Crabbe JC, Adelman JP, et al. A mouse model of episodic ataxia type-1. *Nature neuroscience*. 2003; 6(4):378–83. doi: [10.1038/nn1025](https://doi.org/10.1038/nn1025) PMID: [12612586](https://pubmed.ncbi.nlm.nih.gov/12612586/)
65. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet*. 2009; 5(6):e1000529. doi: [10.1371/journal.pgen.1000529](https://doi.org/10.1371/journal.pgen.1000529) PMID: [19543373](https://pubmed.ncbi.nlm.nih.gov/19543373/)

66. Pugach I, Delfin F, Gunnarsdottir E, Kayser M, Stoneking M. Genome-wide data substantiate Holocene gene flow from India to Australia. *Proc Natl Acad Sci U S A*. 2013; 110(5):1803–8. doi: [1211927110 \[pii\] 10.1073/pnas.1211927110](#) PMID: [23319617](#)
67. Rasmussen M, Guo X, Wang Y, Lohmueller KE, Rasmussen S, Albrechtsen A, et al. An Aboriginal Australian genome reveals separate human dispersals into Asia. *Science*. 2011; 334(6052):94–8. doi: [science.1211177 \[pii\] 10.1126/science.1211177](#) PMID: [21940856](#)
68. Carlsson S, Ahlborn A, Lichtenstein P, Andersson T. Shared genetic influence of BMI, physical activity and type 2 diabetes: a twin study. *Diabetologia*. 2013; 56(5):1031–5. doi: [10.1007/s00125-013-2859-3](#) PMID: [23404445](#)
69. Gunther T, Schmitt AO, Bortfeldt RH, Hinney A, Hebebrand J, Brockmann GA. Where in the genome are significant single nucleotide polymorphisms from genome-wide association studies located? *Omics: a journal of integrative biology*. 2011; 15(7–8):507–12. doi: [10.1089/omi.2010.0154](#)
70. Gruber CJ, Gruber DM, Gruber IM, Wieser F, Huber JC. Anatomy of the estrogen response element. *Trends Endocrinol Metab*. 2004; 15(2):73–8. doi: [10.1016/j.tem.2004.01.008](#) S1043276004000098 [pii] PMID: [15036253](#)
71. Hughes MJ, Liang HM, Jiricny J, Jost JP. Purification and characterization of a protein from HeLa cells that binds with high affinity to the estrogen response element, GGTCAGCGTGACC. *Biochemistry*. 1989; 28(23):9137–42. PMID: [2605247](#)
72. Naciff JM, Hess KA, Overmann GJ, Torontali SM, Carr GJ, Tiesman JP, et al. Gene expression changes induced in the testis by transplacental exposure to high and low doses of 17[alpha]-ethynyl estradiol, genistein, or bisphenol A. *Toxicol Sci*. 2005; 86(2):396–416. doi: [kfi198 \[pii\] 10.1093/toxsci/kfi198](#) PMID: [15901920](#)
73. Yang N, Kazazian HH Jr, L1 retrotransposition is suppressed by endogenously encoded small interfering RNAs in human cultured cells. *Nat Struct Mol Biol*. 2006; 13(9):763–71. doi: [nsmb1141 \[pii\] 10.1038/nsmb1141](#) PMID: [16936727](#)
74. Ikeno M, Suzuki N, Kamiya M, Takahashi Y, Kudoh J, Okazaki T. LINE1 family member is negative regulator of HLA-G expression. *NucleicAcidsRes*. 2012; 40(21):10742–52. doi: [10.1093/nar/gks874. gks874 \[pii\] PMID: 23002136](#)
75. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012; 491(7422):56–65. [http://www.nature.com/nature/journal/v491/n7422/abs/nature11632.html#supplementary-information](#). doi: [10.1038/nature11632](#) PMID: [23128226](#)
76. Nelson SC, Doherty KF, Pugh EW, Romm JM, Ling H, Laurie CA, et al. Imputation-based genomic coverage assessments of current human genotyping arrays. G3 (Bethesda). 2013; 3(10):1795–807. doi: [10.1534/g3.113.007161](#) g3.113.007161 [pii] PMID: [23979933](#)
77. Guo T, Hanson RL, Traurig M, Muller YL, Ma L, Mack J, et al. TCF7L2 is not a major susceptibility gene for type 2 diabetes in Pima Indians: analysis of 3,501 individuals. *Diabetes*. 2007; 56(12):3082–8. doi: [db07-0621 \[pii\] 10.2337/db07-0621](#) PMID: [17909099](#)
78. Replication DIG, Meta-analysis C. Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nat Genet*. 2014; 46(3):234–44. doi: [10.1038/ng.2897](#) PMID: [24509480](#)
79. Coventry A, Bull-Otterson LM, Liu X, Clark AG, Maxwell TJ, Crosby J, et al. Deep resequencing reveals excess rare recent variants consistent with explosive population growth. *Nat Commun*. 2010; 1:131. doi: [ncomms1130 \[pii\] 10.1038/ncomms1130](#) PMID: [21119644](#)
80. Dickson SP, Wang K, Krantz I, Hakonarson H, Goldstein DB. Rare variants create synthetic genome-wide associations. *PLoS Biol*. 2010; 8(1):e1000294. doi: [10.1371/journal.pbio.1000294](#) PMID: [20126254](#)
81. Kramps T, Peter O, Brunner E, Nellen D, Froesch B, Chatterjee S, et al. Wnt/wingless signaling requires BCL9/legless-mediated recruitment of pygopus to the nuclear beta-catenin-TCF complex. *Cell*. 2002; 109(1):47–60. PMID: [11955446](#)
82. Wagner R, Staiger H, Ullrich S, Stefan N, Fritsche A, Haring HU. Untangling the interplay of genetic and metabolic influences on beta-cell function: Examples of potential therapeutic implications involving TCF7L2 and FFAR1. *Molecular metabolism*. 2014; 3(3):261–7. doi: [10.1016/j.molmet.2014.01.001](#) PMID: [24749055](#)
83. Okamoto K, Iwasaki N, Nishimura C, Doi K, Noiri E, Nakamura S, et al. Identification of KCNJ15 as a susceptibility gene in Asian patients with type 2 diabetes mellitus. *Am J Hum Genet*. 2010; 86(1):54–64. doi: [10.1016/j.ajhg.2009.12.009](#) PMID: [20085713](#)
84. Robbins CA, Tempel BL. Kv1.1 and Kv1.2: similar channels, different seizure models. *Epilepsia*. 2012; 53 Suppl 1:134–41. doi: [10.1111/j.1528-1167.2012.03484.x](#) PMID: [22612818](#)

85. Al Safar HS, Cordell HJ, Jafer O, Anderson D, Jamieson SE, Fakiola M, et al. A genome-wide search for type 2 diabetes susceptibility genes in an extended Arab family. *AnnHumGenet.* 2013; 77(6):488–503. doi: [10.1111/ahg.12036](https://doi.org/10.1111/ahg.12036) PMID: [23937595](https://pubmed.ncbi.nlm.nih.gov/23937595/)
86. Rorsman P, Berggren PO, Bokvist K, Ericson H, Mohler H, Ostenson CG, et al. Glucose-inhibition of glucagon secretion involves activation of GABAA-receptor chloride channels. *Nature.* 1989; 341(6239):233–6. doi: [10.1038/341233a0](https://doi.org/10.1038/341233a0) PMID: [2550826](https://pubmed.ncbi.nlm.nih.gov/2550826/)
87. Taneera J, Jin Z, Jin Y, Muhammed SJ, Zhang E, Lang S, et al. gamma-Aminobutyric acid (GABA) signalling in human pancreatic islets is altered in type 2 diabetes. *Diabetologia.* 2012; 55(7):1985–94. doi: [10.1007/s00125-012-2548-7](https://doi.org/10.1007/s00125-012-2548-7) PMID: [22538358](https://pubmed.ncbi.nlm.nih.gov/22538358/)
88. Ho LK, Liu D, Rozycka M, Brown RA, Fry MJ. Identification of four novel human phosphoinositide 3-kinases defines a multi-isoform subfamily. *Biochem Biophys Res Commun.* 1997; 235(1):130–7. doi: [10.1006/bbrc.1997.6747](https://doi.org/10.1006/bbrc.1997.6747) PMID: [9196049](https://pubmed.ncbi.nlm.nih.gov/9196049/)
89. Shia WC, Ku TH, Tsao YM, Hsia CH, Chang YM, Huang CH, et al. Genetic copy number variants in myocardial infarction patients with hyperlipidemia. *BMC Genomics.* 2011; 12 Suppl 3:S23. doi: [10.1186/1471-2164-12-S3-S23](https://doi.org/10.1186/1471-2164-12-S3-S23) PMID: [22369086](https://pubmed.ncbi.nlm.nih.gov/22369086/)
90. Poliak S, Gollan L, Martinez R, Custer A, Einheber S, Salzer JL, et al. Caspr2, a new member of the neurexin superfamily, is localized at the juxtaparanodes of myelinated axons and associates with K⁺ channels. *Neuron.* 1999; 24(4):1037–47. PMID: [10624965](https://pubmed.ncbi.nlm.nih.gov/10624965/)
91. Buchner DA, Geisinger JM, Glazebrook PA, Morgan MG, Spiezio SH, Kaiyala KJ, et al. The juxtaparanodal proteins CNTNAP2 and TAG1 regulate diet-induced obesity. *Mamm Genome.* 2012; 23(7–8):431–42. doi: [10.1007/s00335-012-9400-8](https://doi.org/10.1007/s00335-012-9400-8) PMID: [22961258](https://pubmed.ncbi.nlm.nih.gov/22961258/)
92. Velez Edwards DR, Naj AC, Monda K, North KE, Neuhaus M, Magvanjav O, et al. Gene-environment interactions and obesity traits among postmenopausal African-American and Hispanic women in the Women's Health Initiative SHARe Study. *Hum Genet.* 2013; 132(3):323–36. doi: [10.1007/s00439-012-1246-3](https://doi.org/10.1007/s00439-012-1246-3) PMID: [23192594](https://pubmed.ncbi.nlm.nih.gov/23192594/)
93. Vuillaume ML, Naudion S, Banneau G, Diene G, Cartault A, Cailley D, et al. New candidate loci identified by array-CGH in a cohort of 100 children presenting with syndromic obesity. *Am J Med Genet A.* 2014; 164A(8):1965–75. doi: [10.1002/ajmg.a.36587](https://doi.org/10.1002/ajmg.a.36587) PMID: [24782328](https://pubmed.ncbi.nlm.nih.gov/24782328/)
94. Ritzel MW, Ng AM, Yao SY, Graham K, Loewen SK, Smith KM, et al. Molecular identification and characterization of novel human and mouse concentrative Na⁺-nucleoside cotransporter proteins (hCNT3 and mCNT3) broadly selective for purine and pyrimidine nucleosides (system cib). *J Biol Chem.* 2001; 276(4):2914–27. doi: [10.1074/jbc.M007746200](https://doi.org/10.1074/jbc.M007746200) PMID: [11032837](https://pubmed.ncbi.nlm.nih.gov/11032837/)
95. Loos RJ, Lindgren CM, Li S, Wheeler E, Zhao JH, Prokopenko I, et al. Common variants near MC4R are associated with fat mass, weight and risk of obesity. *Nat Genet.* 2008; 40(6):768–75. doi: [10.1038/ng.140](https://doi.org/10.1038/ng.140) PMID: [18454148](https://pubmed.ncbi.nlm.nih.gov/18454148/)
96. Willer CJ, Speliotes EK, Loos RJ, Li S, Lindgren CM, Heid IM, et al. Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nat Genet.* 2009; 41(1):25–34. doi: [10.1038/ng.287](https://doi.org/10.1038/ng.287) PMID: [19079261](https://pubmed.ncbi.nlm.nih.gov/19079261/)
97. Chambers JC, Elliott P, Zabaneh D, Zhang W, Li Y, Froguel P, et al. Common genetic variation near MC4R is associated with waist circumference and insulin resistance. *Nat Genet.* 2008; 40(6):716–8. doi: [10.1038/ng.156](https://doi.org/10.1038/ng.156) PMID: [18454146](https://pubmed.ncbi.nlm.nih.gov/18454146/)
98. Wen W, Cho YS, Zheng W, Dorajoo R, Kato N, Qi L, et al. Meta-analysis identifies common variants associated with body mass index in east Asians. *Nat Genet.* 2012; 44(3):307–11. doi: [10.1038/ng.1087](https://doi.org/10.1038/ng.1087) PMID: [22344219](https://pubmed.ncbi.nlm.nih.gov/22344219/)
99. Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, Jackson AU, et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet.* 2010; 42(11):937–48. doi: [10.1038/ng.686](https://doi.org/10.1038/ng.686) PMID: [20935630](https://pubmed.ncbi.nlm.nih.gov/20935630/)
100. Guo Y, Lanktree MB, Taylor KC, Hakonarson H, Lange LA, Keating BJ, et al. Gene-centric meta-analyses of 108 912 individuals confirm known body mass index loci and reveal three novel signals. *Hum Mol Genet.* 2013; 22(1):184–201. doi: [10.1093/hmg/dd3396](https://doi.org/10.1093/hmg/dd3396) PMID: [23001569](https://pubmed.ncbi.nlm.nih.gov/23001569/)
101. Xu B, Goulding EH, Zang K, Cepoi D, Cone RD, Jones KR, et al. Brain-derived neurotrophic factor regulates energy balance downstream of melanocortin-4 receptor. *Nature neuroscience.* 2003; 6(7):736–42. doi: [10.1038/nn1073](https://doi.org/10.1038/nn1073) PMID: [12796784](https://pubmed.ncbi.nlm.nih.gov/12796784/)
102. McCaffery JM, Papandonatos GD, Peter I, Huggins GS, Raynor HA, Delahanty LM, et al. Obesity susceptibility loci and dietary intake in the Look AHEAD Trial. *Am J Clin Nutr.* 2012; 95(6):1477–86. doi: [10.3945/ajcn.111.026955](https://doi.org/10.3945/ajcn.111.026955) PMID: [22513296](https://pubmed.ncbi.nlm.nih.gov/22513296/)
103. Murphy R, Carroll RW, Krebs JD. Pathogenesis of the metabolic syndrome: insights from monogenic disorders. *Mediators of inflammation.* 2013; 2013:920214. doi: [10.1155/2013/920214](https://doi.org/10.1155/2013/920214) PMID: [23766565](https://pubmed.ncbi.nlm.nih.gov/23766565/)

104. Skowronski AA, Morabito MV, Mueller BR, Lee S, Hjorth S, Lehmann A, et al. Effects of a novel MC4R agonist on maintenance of reduced body weight in diet-induced obese mice. *Obesity (Silver Spring)*. 2014; 22(5):1287–95. doi: [10.1002/oby.20678](https://doi.org/10.1002/oby.20678) PMID: [24318934](https://pubmed.ncbi.nlm.nih.gov/24318934/)
105. Olney JJ, Sprow GM, Navarro M, Thiele TE. The protective effects of the melanocortin receptor (MCR) agonist, melanotan-II (MTII), against binge-like ethanol drinking are facilitated by deletion of the MC3 receptor in mice. *Neuropeptides*. 2014; 48(1):47–51. doi: [10.1016/j.npep.2013.11.001](https://doi.org/10.1016/j.npep.2013.11.001) PMID: [24290566](https://pubmed.ncbi.nlm.nih.gov/24290566/)